DOCUMENT RESUME

ED 034 185 AL 002 191

AUTHOR	Gleitman, Lila R.
TITLE	Compound Nouns and English Speakers.
SPONS AGENCY	Eastern Pennsylvania Psychiatric Inst., Philadelphia.; National Inst. of Health, Bethesda, Md.
PUB DATE	10 Nov 67
NOTE	207p.
EDRS PRICE	EDRS Price MF-$1.00 HC-$10.45
DESCRIPTORS	Descriptive Linguistics, *English, *Language Research, *Linguistic Competence, Linguistic Performance, *Linguistic Theory, *Nominals, Transformation Generative Grammar
IDENTIFIERS	Grammaticality, *Paraphrasing

ABSTRACT

The author raises the question of how to relate modern transformational grammars to a body of empirical fact, and suggests why paraphrasing might legitimately be considered a feature of behavior relevant to linguistic competence. This study is introduced by a discussion of the empirical basis of descriptive linguistics, followed by sections discussing empirical evidence for grammatical organization and linguistic performance and competence. Section III presents a description of the author's experimental study of the use of compound nouns, involving the free generation of paraphrases, forced-choice, and some replication. The final section discusses the perception of paraphrastic relations in the light of individual differences, grammaticalness and paraphrasability, and semigrammaticalness and error-type. General comments on linguistic method and innovation and creativity in language use conclude the study. Appended are a list of stimuli, tables indicating distribution of errors in the generating task and a new scoring technique, and a bibliography. (AMM)

ED034185

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Compound Nouns and English Speakers[1]

Lila R. Gleitman
Eastern Pennsylvania Psychiatric Institute

Philadelphia, Penna.
November 10, 1967

AL 002 191

---

[1] The work reported here was supported in part by Eastern Pennsylvania Psychiatric Institute, and in part by Grant No. MH 07990 from the National Institutes of Health. I wish to express my appreciation to Elizabeth Shipley and Carlota Smith, who helped in the design and analysis of the experimental work. Thanks also are due to Henry Gleitman, who is largely responsible for the statistical design.

Table of Contents

Introduction: The Empirical Basis of Descriptive Linguistics....1

Section I: Empirical Evidence for Grammatical Organization......15

- A. The ability to classify sentences as grammatical.........20
- B. The perception of phrase-structure and transformational complexity..23
- C. Paraphrase...38

Section II: Performance and Competence.........................43

Section III: An Experimental Study of the Use of Compound Nouns...48

- A. A linguistic description of compound nouns...............49
- B. Empirical studies of the use of compound nouns...........71
  - 1. Pilot work..76
  - 2. Experiment 1: The free generation of paraphrases.....79
  - 3. Experiment 2: Forced-choice..............................123
  - 4. Experiment 3: Some replication...........................139

Section IV: Discussion...142

- A. The perception of paraphrastic relations.................142
  - 1. Individual differences...................................142
  - 2. Grammaticalness and paraphrasability.....................152
  - 3. Semigrammaticalness and error-type.......................165
- B. Some more general comments...............................177
  - 1. Linguistic method..177
  - 2. Innovation and creativity in language use...............185

Appendix A: List of Stimuli.....................................191

Appendix B: Distribution of errors in the generating task, and a new scoring technique...194

Bibliography..197

Introduction: The Empirical Basis of Descriptive Linguistics

> For a small group of specialists, knowing about language is an end in itself. These specialists call themselves _linguists_, and the organized body of information which their investigations produce is called _linguistics_.

This definition of linguists and linguistics, given by Charles Francis Hockett (35) would seem to contain little that could be regarded as controversial. Yet it is by now abundantly clear that there are serious differences among linguists about what is meant when we say 'language', and thus about what the data for linguists' inquiries are or ought to be. Even within the last fifty years there have existed at least three distinguishable approaches to the problem of describing living languages, each with its own methods, each set of methods dictated at least in part by a different view of what the subject-matter was: what, in the world, was being described. Historically, each school has had its moment of dominance: the first, Traditionalist linguistics, the second, Empiricist linguistics, the third, Introspectionist linguistics. I derive these terms, of course, from the general tone of the polemics that accompanied the decline and fall of the first two schools (the third has not yet had time or cause to fall).

Each new generation of linguists has vigorously objected to the methods of its immediate predecessors. It seems to me, however, that in each instance it was _the empirical basis_ of preceding work, even more than the method used to order and codify it, that was crucially being brought into question. Each time the consensus of linguists has swung toward a different methodological approach, the change was accompanied by (and in part attributable to) a shift in the subject-matter under study. To clarify this point, I briefly review here the theoretical stance (as seen both by practitioners and critics) of each of these schools of thought. In each case, the linguist is criticized at least as much for what he chooses to observe as for the way he observes it.

The classical grammarians distinguished a special kind of usage: the 'grammatical' or 'correct' or 'proper' kind of speech and writing. They derived examples of such correct usage, along with the conviction that it existed, from a minute examination of their own and other respectable intuitions. A strong reaction against traditionalist grammars, so devised, came from the Bloomfieldian empiricist school, whose members decried the classicists' preoccupation with 'prestige' dialects. The empiricists believed that the distinction between 'correct' and 'incorrect' speech, drawn by the classicists, was at bottom but a distinction between the dialects of social classes. Thus the traditional grammarian was accused primarily of pettish moralizing. It was said that he described language 'as it ought to be', when he ought to be describing language 'as it was', that he extolled the dialect of the ruling class as though it had some intrinsic linguistic virtue. In fact, it is easy to find in the work of grammarians at the turn of the century and before a perhaps irrelevant emphasis on mere etiquette:

> A peculiar kind of vagabond language, always hanging on the outskirts of legitimate speech, but continually straying or forcing its way into the most respectable company, is what we call slang. The prejudice against this form of speech is to be encouraged....In the first place, all human speech is intended for the ears of others, and must therefore have a certain dignity, a certain courtesy, out of respect to one's hearers if not to one's self. Now slang, from the very fact that it is slang...has a taint of impropriety about it which makes it offensive. Again, the very currency of slang depends upon its allusions to things which are not supposed to be universally familiar or generally respectable; and hence it is vulgar....Finally, the unchecked and habitual use of slang (even polite slang) is deleterious to the mind....
>
> Greenough and Kittredge (27)

As a consequence of this concern with style and nicety, traditional grammars often display a lopsided preoccupation with just those points in the language where class dialects subtly diverge. Yet in fact this cannot sensibly have been the cause of the enormous reaction against traditional grammars by its detractors. Harangue aside, the objection of the empiricists to traditional grammars is to those prescriptive statements that are independent of dialect distinctions, i.e. judgments of inequality concerning utterances even within an idiolect. The real problem the empiricists reacted to was this: on what basis, in fact, can one distinguish 'vagabond' from 'legitimate' speech when the former is admittedly found even 'in the most respectable company'? That is to say, certain expressions, modes of speech, structures, etc., were rejected by classicists even though they had broad currency. The traditional grammarian never could or would say where the empirical authority for his rules came from; the substitute for such authority was too often facile certainty:

> The bewildering variety of our language, and in particular the lawless and fantastic coinages which we have just now been studying, may well suggest the question: 'Is there any criterion of good English? What principle of selection is one to follow who wishes to speak and write his mother tongue with purity and without affectation?' It is the business of grammar, rhetoric, and lexicography to answer this question.
>
> (Greenough and Kittredge, op. cit.)

Among American linguists of the 1930's and 40's it was widely believed that no nonarbitrary empirical basis for the classicists' prescriptions existed.

The empiricists in their turn tried to develop a study of language that had a clearly observable and selfevident data source: obtained speech. Their doctrine was that language would be describe more or less directly as a codification of some large corpus of utterances; that language 'is' no more nor less than a collection of utterances, real and potential. Eugene Nida (1949) below expresses the general consensus of this era:

> The linguist records the actual forms employed, rather than regularizing the data or evaluating utterances on the basis of some literary dialect. In other words,

> it is what people say rather than what some people think they should say that is important to the descriptive linguist....If any judgments are to be passed upon the acceptability or so-called correctness of some usage, these are left either to the anthropologist or sociologist for an objective statement of the factors in the society which make certain persons more socially prominent and hence make their speech more acceptable, or to the man on the street, who is thoroughly accustomed to forming judgments upon the basis of his own egocentric attitudes and limited knowledge.
>
> (65)

A similar program of research, and a similar objection to the traditionalists, is given by Hockett:

> It may come as a shock to learn that the linguist is not particularly interested in...questions [of correctness]...[This] does not mean that the linguist is an advocate of incorrect forms, or that he denies the reality of the distinction between correct and incorrect. As a user of language, the linguist is bound by the conventions of his society just as everyone else is -- and is allowed the same degrees and kinds of freedoms within these conventions....But this has little if any relationship to his special concern, which is analysing language. As an analyst of language, the linguist is bound to observe and record 'incorrect' forms as well as 'correct' ones -- if the language with which he is working makes such a distinction.
>
> (35)

Hockett's concluding comment reveals the position: he doubts even that the distinction between 'correct' and 'incorrect' exists independently of dialect difference, for every language. Bloomfield himself writes:

> The discrimination of elegant or 'correct' speech is a by-product of certain social conditions. The linguist has to observe it as he observes other linguistic phenomena....This is only one of the problems of linguistics and, since it is not a fundamental one, it can be attacked only after many other things are known.

It will have been noticed that the empiricists tended to equate two senses of usage that nowadays are distinguished: there is the question of social (and regional and temporal) dialect difference, for which caveats of correctness are no doubt inappropriate; then there is the question of formal, or written, vs informal styles of expression, where the matter is more complex. Although even in the heyday of scientific linguistics this distinction was not wholly forgotten, the issues of 'acceptability' and 'correctness' as valid perceptual experiences were minimized, if not ignored. The question for linguistics became simply: 'What is said?' Thus did American descriptive linguistics react against the doctrinaire, ill-formulated and, at the time, essentially untestable pronouncements of traditional grammarians. Every utterance was now said to have equal status as a linguistic datum, the fact of having-been-uttered the definition of grammaticalness.

The most serious effort to follow these principles in writing a grammar was made by Charles Fries (2.4). He tapped a large number of telephone conversations in the belief that language could be described by beginning with a corpus of spontaneous speech. From these data there were eventually distilled class-lists equivalent in intent to exhaustive co-occurrence matrices, and corresponding very grossly to the traditional parts of speech. A great step was thereby taken: the new grammars described the regularities of language rather than emphasizing 'exceptions' and stylistic minutiae; the episodic character of classical grammars was counteracted.

Various questions remain, however. It has been objected that 'intuition' is not evaded by such a method, i.e. that the lists of classes were not mechanically derivable from the raw data, as these linguists supposed. Far more seriously, it has been objected that the resulting grammars were not 'interesting', 'comprehensive', 'illuminating', etc. It is a curious philosophical excess of the empiricist movement that this last kind of criticism could, as a kind of literary conceit, be accepted as a compliment:

> Children want explanation, and there is a child in each of us; descriptivism makes a virtue of not pampering that child.

So wrote Martin Joos in 1958 (40), justifying the replacement of phonological explanations by 'a sober taxonomy'. Then in this revolutionary period we find a clear gain in justifiability and freedom from arbitrariness but, some say, a comparable loss in compactness and 'explanatory power'.

During the Bloomfieldian period, there was at least one linguist of stature who remained aloof from prevailing opinion concerning the study of language. This, of course, was Edward Sapir, who was suggesting a new frame of reference for descriptive linguistics. His overriding interest was in the perceptual organization that gives rise to overt language behavior -- in fact, just that description of 'underlying mental processes' that Bloomfield was at pains to banish. In those days of aggressively scientific linguistics (matched rather neatly in neighboring psychological circles by mechanistic behaviorism), Sapir's views were regarded as somewhat subversive -- a dangerous backsliding. Martin Joos, again, summing up Sapir's contribution, wrote:

> [Sapir's] contribution was not the developing of any method, but rather the establishing of a charter for the free intellectual play of personalities more or less akin to his own. If their wits happen to be dimmer (and here he had few equals), their blunders may betray the essential irresponsibility of what has been called Sapir's 'method'. We welcome the insights of his genius, which allowed no scrap of evidence to escape at least subconscious weighing; where it is possible to check up, we normally find him right; thus we seem captious when we point out that he also said many things which are essentially uncheckable ('invulnerable') and thus not science.

Ten years later, the views of Sapir do not sound so odd or mystical. He argued that a really explanatory theory of language would be found in a higher-level construct of mental patterning; his view is exemplified in this discussion of phonetics:

> I found it difficult or impossible to teach an Indian to make phonetic distinctions that did not correspond to 'points in the pattern of his language', however these differences might strike our objective ear, but that subtle, barely audible, phonetic differences, if only they hit the 'points in the pattern', were easily and voluntarily expressed in writing. In watching my Nootka interpreter write his language, I often had the curious feeling that he was transcribing an ideal flow of phonetic elements which he had heard, inadequately from a purely objective standpoint, as the intention of the actual rumble of speech. (68)

The difference between Sapir's and Bloomfield's 'methods' was not in the recognition of phonetic patterning (for phonology, at least, the difference between physical and perceptual fact was acknowledged), but in Sapir's "mentalistic" assumption that the perceptual systematization at times transcended present acoustic cues -- that the speaker heard what wasn't there, that he had internalized and understood much that he had never heard. A most beguiling example reported by Sapir is recorded by Bernard Bloch in an editorial note (3):

> Sapir was working once with a bilingual Navaho informant, an old woman who patiently recited paradigms for him. When he had heard several dozen forms, Sapir ventured to create new forms by extrapolation. The old woman shook her head in wonder, and looked at him respectfully. 'You're a funny man,' she said, 'You say things in my own language that I never heard before, and they're right!'

In 1957 Noam Chomsky carried Sapir's position to its implied conclusion: Chomsky outlined a new linguistics which had as an avowed aim the description of the speakers' intuitions concerning the forms of language. This represented a radical departure from the strict equalitarian attitude toward utterances that by now had become the establishment view within the field, and an angry dialogue has persisted until today. Chomsky claimed not that the speaker could reliably verbalize rules or regularities of his native tongue, but rather that the speaker could make judgments and construct utterances that were explicable only in terms of an implicit systematization of the language. Although Chomsky writes that a 'corpus' of utterances underlies his own linguistic theory construction, it should not be supposed that what is meant is the kind of corpus the empiricists thought they were using. Chomsky writes:

> In order to set the aims of grammar significantly it is sufficient to assume a partial knowledge of sentences and nonsentences....For the purposes of this discussion...suppose that we assume intuitive knowledge of the grammatical sentences of English and ask what sort of grammar will be able to do the job of producing these in some effective and illuminating way. We thus face a familiar task of explication of some intuitive concept, in this case, the concept 'grammatical in English,' and, more generally, the concept 'grammatical.'
> (14)

The corpus (and thus the subject-matter) is now redefined: While the empiricist uses 'some obtained utterances', Chomsky selects some utterances judged to be grammatical and some (obtained or not, indifferently) utterances judged not to be grammatical.

In the context of the present discussion, it is important to notice that 'correctness', 'acceptability', and 'grammaticalness' -- far from being excluded as secondary or irrelevant -- become in the theories of Sapir and Chomsky central concepts. Thus the question has shifted from what the grammarian thinks the speaker ought to say, toward what the speaker says, to what the speaker thinks he ought to say; from what should by rights be said to him toward what is actually said to him to what he hears.

Much of the attack and counterattack in the scholastic war that has ensued seems to converge around transformations as a formal device, but it is easy to see that the struggle pivots around Chomsky and his group. Zellig Harris, who, after all, introduced the concept of transformations (sentential relations) into syntactic study, was not engaged in the combat. Harris' work was not vitally controversial: in his published work on transformations (28, 29) he appears to tap the same empirical sources as his predecessors. Chomsky, on the other hand, draws from a hypothetical set of nice judgments by speakers.

Many American linguists persist in judging transformational grammars by their observational accuracy in describing speech (see, e.g., Hill (32)). If such were the intention of generative grammars, or of any transformational grammars, then surely they fail to meet the grossest test. But a theory cannot sensibly be faulted by reference to a set of data it never set out to pacify. Nevertheless, the assumption of classicists that the empirical issues could be passed over by edict seems untenable. Traditionalists and transformationalists can make common cause in the concern with judgmental features of linguistic behavior, but the need for external validation can hardly be questioned.

The dispute between empiricist and transformational linguists involves a disagreement concerning the very subject-matter of linguistics. The empiricist contention that language can be described simply by classifying and arranging the elements of overt speech seems inadequate. This argument has been made so well, so frequently, and so vociferously in the literature (14, 15, 66) that it is superfluous here to add a new version. Yet the central question the empiricists raise requires a response: In what nonarbitrary manner are certain utterances removed from consideration? In other words, what are the empirical correlates of transformational linguistics? So long as this question is left unanswered, transformational grammars remain descriptions in search of a phenomenon.

We therefore raise here the question of how to relate modern transformational grammars to a body of empirical fact; similarly, the question of what aspects of reality will serve to justify or validate these grammars, aspects of these grammars, details of these grammars, most important, relevance of these grammars to the totality of the study of language.

A systematic organization of language, along the lines of a generative and transformational grammar, impinges on the activities of a human being when he speaks or comprehends his native tongue. This matter is hardly in doubt. I am concerned here only with the place of such grammars in a description of language and language use. In the empirical work described here, I examine certain rockbottom factual assumptions of the transformational grammarians, and describe some tests of the adequacy of these assumptions. From such work I believe something can be contributed to the understanding of the methodological bases of current linguistic research, and to the question of where transformational grammars fit into a description of language.

A note: Earlier I quoted from writers who felt that issues such as the perception of 'correctness' lay outside the linguists' competence and rightful concern. They are clearly in again. It is only one more argument that the borders between sciences are not sensibly subject to juridical decision. Hockett's definition of linguists and linguistics is ultimately the only tenable one, for it is the only one vague enough, for scientific purposes; linguistics is what linguists do. It is perhaps unfortunate that Chomsky's recognition of covert aspects of linguistic competence has opened a Pandora's Box of psychological and psychophysical problems. His own earlier assumption was that 'simple tests of acceptability' could be designed to validate transformational grammars; it has not happened so. Outside the cloisters of the grammarians there are masses of people who seem as little able to make judgments as to speak correctly, as little able to display their 'competence' as to perform stably. If we must therefore enter into a period of empirical investigation, some traditional academic boundary lines will surely be overrun. Psychologists have contributed little enough to the study of language that we need surely have no qualms at invading their domain. I raise the issue here because expectably little derives from this work to controvert Chomsky's position: that transformational grammars are ample and able to problems of syntax and semantics; but I will suggest that the place of syntax and semantics in the study of language needs clarification. This may be cause for argument, but not for declassification, or even reclassification, of either linguists or linguistics. Perhaps especially in the study of language, we have to accept a continual overlapping with all the sciences of man. Sapir is here again instructive:

> It is clear that the interest in language has in recent years been transcending the strictly linguistic circles. This is invevitable, for an understanding of language mechanisms is necessary for the study of both historical problems and problems of human behavior. One can only hope that linguists will become increasingly aware of the significance of their subject in the general field of science and will not stand aloof behind a tradition that threatens to become scholastic when not vitalized by interests which lie beyond the formal interest in language itself.... (69)

Section I: A Historical Review: The Empirical Evidence for Grammatical Organization

I have suggested that the choice of subject-matter schismatizes linguists into 'schools' even before the problem of treatment of the data is raised. While the subject-matter of American empirical linguistics (speech) is relatively easy to define and justify, the fragment of reality described by transformational (and traditional) grammarians is not altogether clear.

In 1957, in Syntactic Structures, Chomsky set as the fundamental aim of grammars the description of projective features of language, i.e. of an unbounded set of grammatical expressions. The 'creative aspect' of language was viewed as a psychological as well as a formal reality; the ability of speakers to form indefinitely many novel sentences was taken as the basic phenomenon requiring descriptive counterpoint in a grammar:

> The fundamental aim in the linguistic analysis of a language L is to separate the grammatical sequences which are the sentences of L from the ungrammatical sequences which are not sentences of L and to study the structure of grammatical sequences.
>
> A grammar mirrors the behavior of the speaker, who, on the basis of a finite and accidental experience with language, can produce or understand an indefinite number of new sentences. Indeed any explication of the notion 'grammatical in L' (i.e., any characterization of 'grammatical in L' in terms of 'observed utterance of L') can be thought of as offering an explanation of this fundamental feature of linguistic behavior.

If it is the case that humans can say and understand new utterances, failure to account for this feature of linguistic behavior, at least in a psychological, or 'mentalist', grammar of the Chomsky variety, can by itself be taken as proof of the inadequacy of a proposed grammar:

> [Empiricist] speculations have not provided any way to account for or even to express the fundamental fact about the normal use of language, namely the speaker's ability to produce and understand instantly new sentences that are not similar to those previously heard in any physically defined sense or in terms of any notion of frames or classes of elements, nor associated with those previously heard by conditioning, nor obtainable from them by any sort of 'generalization' known to psychology or philosophy.
>
> Chomsky (Aspects of the Theory of Syntax, 1965)

There is in fact no doubt that novel utterances can be elicited even from very young children. On the other hand, Pavlov's dog will salivate to previously unheard bells and the energetic creativity of this latter act is open to some question. The special assertion of Chomsky concerning language is that no physical dimension can be found to account for the production or comprehension of new sentences that are in certain ways 'similar to' previously encountered sequences. It follows, in Chomsky's view, that humans are 'prewired' to recognize certain kinds of linguistic relations and not others. A grammar is then a neurophysiological hypothesis based on an examination of certain response-types.

It should be noticed that even for animal behavior it is often no simple task to specify a physical dimension defining similarity (a rat will respond more emphatically to a tone an octave away from his accustomed tone than to one but a few notes away (2), even though pure tones are used to minimize the chance of audible overtones). 'Similarity' even here is a hindsight notion, dependent on identity of response type. The perception of octaves must be explained neurophysiologically for the animal, i.e. the dimensions of physical similarity that are valid and immediate for ratkind are tied up with intrinsic cognitive functioning. So, for man, the tendency to notice certain kinds of patterning and not others must be at least in part neurophysiologically determined. To this extent certain organizational contours of language learning and use are pregiven in the organism. But it must be recognized that this general fact does not come as a shock to the (post 1930) psychological community, nor does it preclude the role of learning in language, nor does it by itself point to some transcendental feature of human cognition that is in principle irrelevant to the description of other forms of life. However, as we shall later discuss, the existence of anatomic and physiologic correlates of verbal behavior (for one thing) make it extremely unlikely that attempts to describe human language and cognition as mere extensions of animal cognitive functioning will succeed. In fact, disillusionment with the attempt to extrapolate complex thinking processes from models of animal behavior accounts for much of the interest among psychologists in Chomsky's hypotheses.

What is not selfevident about the assertion that speakers can deal with novelty is the generality of this skill. Even in a loose qualitative way we know nothing of the limits, extent, pervasiveness over the linguistic community, or conditions for acquisition and use of novel linguistic behavior. Then though it is factually incontrovertable that certain projective features must be written into language descriptions, it is by no means clear which (and how far and to whom and when) productive devices of various sorts are available. It has been the Chomskian view that any restrictions on these manipulations are only the general factors limiting all sorts of mechanical and cognitive activities (e.g. limitations of memory span, degree of attentiveness, motivation, interference).

A good deal of experimental ingenuity would be required to design a compact and feasible test of the extent of the ability to comprehend and produce sentences, for it is obvious that limitless outpourings of new sequences do not necessarily tap the limits of the structuring postulated in a grammar. Not only must all structural types be sampled, but the relations among sentences and sentential types must be examined; for it is the interweave among structures that lies at the heart of the transformational hypothesis.

No one has ever undertaken to elicit sentences from individuals in a way systematic enough to point to an effective procedure for producing or understanding 'all grammatical' sequences (even within the constraints of memory, etc.). No such heroic procedure is necessary or useful, since the fabric of transformational theory is sufficiently rich and detailed to suggest a host of feasible corroborative tests. Since psychological validity is regarded as the central external buttress supporting proposed grammars, various evidences from behavior are adduced to justify these grammars both in general format and in detail. In this section I will consider the kinds of psychological evidence that have been brought forth to show the adequacy of this approach to grammar. In each instance, I attempt to show why such empirical evidence would support transformational theory, and examine also the results of attempts to make this evidence plausible, i.e. the fate of attempts to validate these assertions experimentally.

The major empirical claims of the transformational grammarians have been: (A) that speakers can judge between grammatical and ungrammatical sequences of English words; (B) that formal definitions of phrase-structure and transformational organization and complexity are mirrored in behavior; (C) that language acquisition is more-or-less uniform in the human species, i.e. it is acquired independently of level of intelligence and of specific environmental features; and (D) that specific paraphrastic (meaning-preserving) relations among sentential types are known to the speaker of the language.

A. The ability to classify sentences as grammatical: A central aspect of transformational grammars, and indeed all grammars, is the specification of all the 'legitimate' sequences in the language. If the grammar mirrors human behavior in this respect, then within certain statable limitations, e.g. length, a speaker of the language ought to recognize an illegitimate sequence when he hears it. In effect, the speaker would realize that his 'internalized grammar' does not specify this sequence. Chomsky suggests a test of this ability in Syntactic Structures:

> One way to test the adequacy of a grammar proposed for [a language] is to determine whether or not the sentences that it generates are actually grammatical, i.e., acceptable to a native speaker, etc. We can take certain steps toward providing a behavioral criterion for grammaticalness so that this test of adequacy can be carried out...
> (14)

Until fairly recently it was assumed by most grammarians of Chomsky's general persuasion that such an empirical test would so selfevidently succeed that there was little reason for systematic experimental investigation. Formal models of language are constructed with this datum accepted as proven. Hilary Putnam, a philosopher, speaking before the American

Mathematical Society in 1960 (67), showed that this hypothetical ability, "makes it legitimate to seek recursive function-theoretic structures which could serve as models for grammars." These are the facts given:

> ...that speakers can presumably classify sentences as acceptable or unacceptable, deviant or nondeviant, et cetera, without reliance on extra-linguistic cues. There are of course exceptions to this rule, but I am more impressed with the multiplicity of nonexceptions. I imagine, for example, that if I were on any number of occasions presented with a list of sentences and asked to say which ones I thought were grammatical, I would on each occasion and without any information on the supposed context of use of the individual sentences classify 'Mary goed home' as an ungrammatical sentence and 'Mary went home' as a grammatical sentence. This act of classifying sentences as grammatical or ungrammatical seems to be one I can perform given no input except the sentences themselves. In short, it seems that in doing this job of classifying, I am implicitly relying on something like an effective procedure.

Empirical investigation of this topic has been attempted. Critics of the general approach were the first to set it to the test, but these investigations were carried out with such astonishing experimental naivity as to render the results inconclusive. Notice that Chomsky indicated that some 'behavioral criterion' could perhaps be developed for testing the speaker's ability to make these judgments. The assumption that classification could be performed by lay subjects in the strictly schematic version outlined by, e.g., Putnam, is obviously too strong. Consider merely the difficulty of knowing in what way the subject understands the instructions, and the complexity of actual experimental verification begins to be obvious.

Nonetheless, rediscovering psychological methodology from its beginnings, linguists such as A. A. Hill (32) submitted to subjects lists of sentences, asking which were acceptable, or grammatical. The results, as expected, are far short of reliable. There seems to be evidence of highly idiosyncratic behavior. However it is almost impossible to interpret these results because the instructions were so vague.

Sleator and Maclay (52) attempted a more systematic study of the same question. In this version, an attempt is made to control for the effects of 'nonsensicalness' and to separate these from the effects of 'grammaticalness'. Nonetheless, there was at minimum a good deal of statistical smear in the outcome: although there was a discernible indication of preference for grammatical sequences, the result was either imperfect, or else masked by extraneous effects of the testing conditions.

Hill concludes that speakers cannot judge sentences in isolation. Sleator and Maclay conclude that empirical criteria cannot be, because they do not here succeed, used to justify grammars. Both conclusions are highly arbitrary. Failure to provide an existence proof in a given experimental situation can scarcely be regarded as disconfirmation; failure of an empirical technique to elicit the desired result -- or the opposite -- hardly justifies giving up investigation. Within linguistics, it was left to Chomsky to point out that, in all human activity, there are deficits in performance that may be wholly extraneous to the topic under experimental review. Here it suffices to say that 'acceptability tests' in this form give garbled results, probably for any of a number of external reasons.

B. The perception of phrase-structure and transformational complexity: Where 'acceptability tests' failed to provide reliable evidence for grammatical structuring, more sophisticated psychological techniques have provided more revealing data.

George Miller and his associates (Miller, 1962; Miller and Isard, 1963) set out to study the relevance of syntactic structure to perception and recall. They presented subjects with grammatically and semantically acceptable sentences, sentences that were grammatical but semantically anomalous, and sentences that were both ungrammatically and semantically anomalous. They then asked subjects to listen to these sentences in the presence of a masking noise. Later, (Marks and Miller, 1964) they asked subjects to memorize the sentences. In both experiments, the results are the same: the grammatical and semantically sensible sentences are most easily dealt with; the scrambled semantically anomalous sentences are the most difficult; the other case falls in between.

Lest it seem obvious or trite to have performed such experiments at all, it should be recalled that even in linguistics today there are those who question the assertion of linguistic organization in the individual speaker, while recognizing regularities in the language community. Even more, in psychology, there have been many attempts to describe language using only the most primitive grammatical and statistical organizing hypotheses. Miller's series of experiments established beyond further doubt that word-sequences with certain kinds of grammatical and semantic properties are perceived wholly differently from haphazard and semistructured sequences. There remains to provide the detail of the organization, and the psychological processes this organization implies.

Bever and Fodor (1963) showed that speakers can recognize internal features of organization (phrase-structure) within the sentence. They interpolated a clicking noise during the recorded presentation of a sentence, and asked subjects to judge at what point in the sentence the click occurred. Judgments tended to be displaced toward major constituent boundaries. In a crucial test, syntactically ambiguous sentences whose alternate interpretations had different major constituent boundaries were presented after the subject was 'set' by prior sentential context to hear one interpretation or the other. Although the stimuli were physically identical the displacement was predictable from the set.

In most transformational grammars, it is possible to distinguish on formal grounds between 'simpler' and 'more complex' constructions. If the grammar mirrors behavior in this sense, the individual will find the simpler construction more immediate and comprehensible than the complex.[1] Miller (1964) reviews a series of attempts to study the relative comprehensibility of affirmative (grammatically simpler) and negative (more complex) sentences, when semantic content is controlled. Briefly summarizing this work: Smoke (1932) and Hovland and Weiss (1953) showed under a variety of experimental circumstances that negative sentences are more difficult to deal with than affirmative sentences. Wason (1961) showed that the affirmative-negative difference was more important than the difference between truth and falsehood, thus suggesting that a grammatical factor beyond the conceptual, or logical, factor accounted for some of the difficulty with negation. McMahon (1963) showed that this difficulty could in part be described as a measurable difference in time required to unscramble the transform. This last explanation depends, of course, on the assumption that sentences must be transformed to -- or understood in -- their simplest syntactic form as a step in comprehension.

---

[1] Such effects are noticeable only when a way is found to tax the individual because, presumably, the language patterns are so well established in an adult that all 'normal' sentence types will be grasped very easily. There are any number of ways (e.g. masking noise, memory loading, interpolated tasks) to make difficult perceptions and activities that are so easy under normal circumstances as to make any error or even delay unlikely.

The general attempt to measure the times involved in 'making grammatical transformations' was begun by Miller (1962). In the early experiments, subjects were shown by example the kind of transformation they were to effect, and the time to response was measured. Rough estimates of the time involved suggested that the effects of complexity were additive (i.e., transforming a negative-passive to a simple active positive took about the time of transforming a negative to a positive plus the time of transforming a passive to an active). It should be remembered, however, that such conscious manipulation of syntactic forms may be an activity quite alien from the normal use of language. In so far as it may, however, be relevant, the result seems consistent with formal definitions of syntactic complexity.

A more natural and powerful technique for measuring relative complexity among syntactic forms has been developed by Savin and Perchonock (70) in the context of studies of immediate-memory. Savin requires subjects to recall a sentence, and in addition to the sentence, as many as possible of an arbitrary list of unrelated words or digits. The number of arbitrary words that are remembered varies inversely with the complexity of the sentence that is to be recalled; again, the relationship seems to be additive. The rationale of these experiments is this: it is presumed that the sentence 'takes up space' in a small 'immediate-memory bank'; a certain amount of space is now left over and can be 'loaded' with the arbitrary material. The simpler the sentence, the less space it should have taken in the memory bank, and the more space should be left for the filler words. In fact, assertions of a generative grammar concerning relative complexity of structural types correctly predict differences in the number of filler words that are recalled by the subjects.

The work summarized above admittedly is relevant only to the more gross aspects of English syntactic structure; the simplest and most general transformational relationships are dealt with under certain very restricted circumstances. The overall result is a display of the fruitfulness of grammatical patterning as a concept in describing linguistic behavior.

C. Faculte de langage: Man is innately endowed with the ability and the propensity to learn to speak. No other animal is so endowed. That language ability is species-specific has rarely been seriously doubted. That the ability to speak is a cognitive function qualitatively different from anything available in the repertoires of other animals is a common supposition. Occasionally there have appeared revolutionary thinkers, such as B. F. Skinner (14), who have tried to construct a psychologically tenable theory of language by extrapolation from the laws of animal thought. However, the existence of anatomic and physiologic correlates of language (i.e., morphological and functional specializations such as cerebral dominance and special coordination centers for motor speech) make it unlikely in the extreme that this faculty of mind will be explained as a mere elaboration of functions available to the lower animals. Attempts to teach animals the rudiments of speech are uniform failures (31, 44). On the other side of this coin, severe physical handicaps such as congenital deafness or blindness, damage to the articulatory mechanisms, even feeble-mindedness, fail to suppress or even significantly to retard the acquisition of language in humans.[1]

In recent years, Chomsky and his coworkers have argued that the 'faculte de langage' represents a large instinctual pattern of predispositions and behaviors in mankind, relatively immune to environmental or cultural control or direction, and more-or-less independent of other aspects of cognition:

> ...the structure of particular languages may very well be largely determined by factors over which the individual has no conscious control and concerning which society may have little choice or freedom. On the basis of the best information now available, it seems reasonable to suppose that a child cannot help constructing a particular sort of transformational grammar to account for the data presented to him, any more than he can control his perception of solid objects or his attention to line and angle. Thus it may well be that the general features of language structure reflect, not so much the course of one's experience, but rather the general character of one's capacity to acquire knowledge -- in the traditional sense, one's innate ideas and innate principles. It seems to me that the problem of clarifying this issue and sharpening our understanding of its many facets provides the most interesting and important reason for the study of descriptively adequate grammars and, beyond this, the formulation and justification of a general linguistic theory that meets the condition of explanatory adequacy....
>
> (Chomsky, 1965)

---

[1] In these points, we are following Lenneberg (1964), who presents current evidence concerning the biological bases of language ability.

Some linguists have expressed mystification at the Chomskian interest in the biological bases of linguistic ability. However, once it is accepted that linguistic theory is a description of the verbal activities and capacities of people rather than a formalism responsible only to its own elegance and consistency, it follows that language acquisition becomes at least coequal with other issues of verbal behavior as a phenomenon that linguistics will have to explain; more precisely, linguistic theory must be able to support a plausible description of language acquisition, though perhaps remaining neutral in the question of the processes and neural structures involved in learning.

The preeminence of language learning as a crux of linguistic theory comes about in the search for 'linguistic universals', hypothesized features of structure that are necessary and fundamental to natural language, hence, possibly, 'innate' organizing principles rooted in the cognitive functioning of human beings:

> A theory of linguistic structure that aims for explanatory adequacy incorporates an account of linguistic universals, and it attributes tacit knowledge of these universals to the child. It proposes, then, that the child approaches the data with the presumption that they are drawn from a language of a certain antecedently well-defined type, his problem being to determine which of the (humanly) possible languages is that of the community in which he is placed. Language learning would be impossible unless this were the case. The important question is: What are the initial assumptions concerning the nature of language that the child brings

> to language learning, and how detailed and specific is the innate schema (the general definition of "grammar") that gradually becomes more explicit and differentiated as the child learns the language? For the present we cannot come at all close to making a hypothesis about innate schemata that is rich, detailed and specific enough to account for the fact of language acquisition. Consequently, the main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning.
>
> (Chomsky, 1965, op.cit.)

It would be beyond the scope of our discussion to relate in detail the form of the language-acquiring-device that Chomsky postulates. But the kind of phenomena and theoretical hypotheses that lead him to the innatist position are well characterized in the following:

> It seems plain that language acquisition is based on the child's discovery of what from a formal point of view is a deep and abstract theory -- a generative grammar of his language -- many of the concepts and principles of which are only remotely related to experience by long and intricate chains of unconscious quasi-inferential steps. A consideration of the character of the grammar that is acquired, the degenerate quality and narrowly limited extent of the available data, the striking uniformity of the resulting grammars, and their independence of intelligence, motivation, and emotional state, over wide ranges of variation, leave little hope that much of the structure of the language can be learned by an organism initially uninformed as to its general character.
>
> (Chomsky, 1965, op. cit.)

Chomsky then is hypothesizing very definite neurophysiologically determined organizing principles for discovering and selecting among grammars. The claim is for a faculty of mind more-or-less independent of the principles and processes that are thought to underly thinking and that are related to 'intelligence' in a more general sense. The human is viewed as a structure-seeking organism; linguistic behavior as a speciated instinct, in much the sense of the animal behavior patterns studied by the ethologists.

The following kinds of evidence would seem to be relevant to testing the correctness, or at least the utility, of this view: (1) the existence of anatomic and physiologic correlates of language behavior; (2) 'critical periods' for language-learning, or certain stages of language-learning, and, in general, a regular developmental onset; and (3) asymptotic similarity in the adult behavior of nonpathological individuals.

1. anatomic and physiologic correlates: Specific areas of the brain basic to speech have been isolated although the relation of these areas to each other and their specific functions are not yet well known. Lenneberg (op. cit.) lists as examples of the specific biological endowments of man necessary to speech: cerebral dominance, specialization of cerebrocortical topography, special coordination centers for motor speech, specialized temporal pattern perception, special respiratory adjustment and tolerance for prolonged speech.

The existence of these special biological endowments is the more suggestive because it implies some separation of linguistic and other intellective functions. The literature on asphasia and other congenital and traumatic brain pathologies also (26; 38; 71) suggests that various cognitive functions may remain unimpaired although language facility has been severely damaged.

2. the onset of language: critical periods: A major factor in instinctual patterns as described by ethologists is the 'critical period' during which the behavior is catalyzed by aspects of the external environment which are ineffectual at other maturational periods. The behavior emerges only when a certain stage of neurophysiological maturation is reached. If the external environment is sufficiently abnormal to preclude the emergence of these patterns at the maturationally appropriate time, the behavior *never* appears in the organism at all. Thus, the study of language development in feral children would be of critical interest in this connection. Roger Brown (10) has reviewed the rather vague literature on this subject, but with the information available no conclusive statements are possible. In general, it seems practically impossible to teach language to those feral children who were outside human society between the ages of one and six; yet there is evidence that those children who were cast out or 'lost' were abnormal in the first place.

On the other hand there are fairly well-documented cases of children who were isolated from human contact for the normal period of language development without permanant linguistic impairment. For example, Davis reports on the case of a child who was tied to a chair and maintained in silent isolation for over six years. This child was able to achieve normal speech (and a high I.Q. score) only two years after rehabilitation (18).

On the basis of studies of language development in the normal child, (1, 7, 11, 12, 19, 41, 58), which are now proliferating, Chomsky and some other transformational linguists are concluding that the onset of language is sufficiently regular to support the maturational hypothesis. Below are remarks from Eric Lenneberg, in support of a biologically determined description of language:

> The onset of speech is an extremely regular phenomenon, appearing at a certain time in the child's physical development and following a fixed sequence of events, as if all children followed the same general "strategy" from the time they begin to the period at which they have mastered the art of speaking...
>
> (Lenneberg, op. cit.)

Lenneberg at this point cites some general aspects of the learning process (e.g., 'the first things that are learned are principles -- not items') but the precise evidence is unclear from his report, although he cites a number of studies in a general way. From our own work on language development (Shipley et al. 1965), the conclusion of the grammarians on this issue seems questionable. Our subjects appear to show variation in onset, and seem to display varying "strategies" of acquisition. It may be that this variation is 'small' in the context of a lifetime, or in the context of abstractly possible variation; but it does not seem to me that there is as yet any basis at all on which notions like 'small and expectable variation' vs 'huge and astonishing variation' could sensibly be defined.

Yet so well entrenched is the belief in a regular and environment-proof onset in development that Chomsky and Lenneberg have seen fit to cite it as an argument for innate endowments. Some of the (very paltry, to be sure) evidence that is now in our possession suggests, on the contrary, that environment is a powerful factor at least in affecting developmental rate (63, concerning the retarding effect of impersonal handling and lack of linguistic contact in an orphanage situation; 37, concerning the accelerating effects on development of story-reading in lower-class environments; 53, concerning the low but consistent correlation of language onset with IQ; etc.).

3. asympototic equivalence in adult linguistic competence: A commonplace among the linguists and philosophers is that, in non pathological individuals, adult linguistic competence is uniform. Given the enormous variation of intelligence, environment, and culture, a finding of adult equivalence would give credence to an innatist hypothesis. For example, Katz (42) writes:

> Variation in performance with intelligence [in non-linguistic tasks] contrasts with the performance of speakers with respect to some purely linguistic skill, where no significant individual differences are found...

In a similar vein, Hilary Putnam asserts the independence of language ability from intelligence as an argument for 'a model of [the speaker] as a Turing machine who is processing each new sentence with which he is provided according to some mechanical program.':

> Even a person of very low-grade intelligence normally learns both to speak his particular dialect grammatically and to recognize deviations from grammaticalness...a moron whose parents happen to speak the prestige dialect may have serious vocabulary deficiencies but he rarely has grammar deficiencies. He too learns to speak the prestige dialect, and to feel that there is something wrong with sentences that deviate from the grammatical regularities of the prestige dialect, even if he does not have the extremely complicated skill (parsing) which is required to say what is wrong. But an ability of this kind, which can be acquired by practically anyone or which can be utilized by practically anyone independently of intelligence level, is almost certainly quasi-mechanical in character.
>
> (Putnam, op. cit.)

Experimental evidence for this view of uniform adult linguistic competence is extremely thin. In tests of 'acceptability' (cf. Hill; Sleator and Maclay), reading comprehension (53), and other similar skills presumably related to language facility, we do not find uniformity at all. It is argued by Chomsky that external factors such as performance decrements and differences in cognitive abilities account for any apparent difference in underlying competence. The question of competence vs performance will be taken up in a later section (Sec. II); here it suffices to say that the notion of uniform competence is not, to my knowledge, supported by any conclusive systematic evidence. That a tremendous degree of systematic organization of linguistic data is a crucial element in the language perception and use of all normal speakers is well established, particularly by the studies of Miller and his coworkers (op. cit.); but, as is obvious, such results are compatible both with hypotheses of individual difference or hypotheses of equality in degree of underlying competence.

An argument often cited in favor of the view that all adults, over a broad spectrum of intellectual ability, come to a similar level of competence is the appearance of speech in the feeble-minded. Lenneberg (op.cit.) writes: 'Children whose IQ is 50 at age 12 and about 30 at age 20 are completely in possession of language although their articulation may be poor and an occasional grammatical mistake may occur.' Thus as measured IQ decreases with age, something like normal speech nevertheless develops.

In the main, the evidence for equal competence is anecdotal (anyone can and does read the Daily News, ergo everyone is linguistically competent). Anecdotal evidence for inequality in linguistic competence is even easier to provide, and it is just as unconvincing (only the very best people can read Dylan Thomas, or make a good pun, or write a grammar, and therefore some are more linguistically competent than others). Even where systematic data are cited (Lenneberg, above), the measure defining 'complete possession of language' is left totally unstated. In the development of transformational grammars, in all their variety and richness, we must make use of the most exquisite judgments of the sentences and semisentences of the language. It is claimed that such judgmental precision is necessary in describing languages, and that the possibility of fine judgments is itself evidence of the remarkable competence of the speaker. Were such delicate decisions excluded in the process of writing or justifying a grammar,[1] grammars could not sensibly be devised or tested nor would there be as much reason to exault the complexity of syntax; for this complexity would not be discovered. But the mere production of a few, or a few thousand, dull but 'previously unheard' sentences by the average speaker is taken as sufficient reason to endow him with all the wealth of the English language. Presumably the mere fact that a linguist's butcher can say "Good morning, Dr. Chomsky; the liver is fine today." serves as proof that in every butcher there have emerged the subtlest features of English syntactic structure. It seems to me selfevident that the matter cannot rest on such flimsy evidence.

---

[1] Those structural linguists who have no grammatical theory to account for these judgments do, indeed, exclude them on the grounds that they are 'not grammar', and the transformationalists are quick to point out the emptiness of this protest.

D. Paraphrase: The ability 'to say it in one's own words' is intuitively taken as an index of understanding. Some grammatical transforms are simple paraphrases; the surface structure of the sentence or phrase is changed, but its deep structure is unchanged; the information conveyed may intuitively seem to be the same. Transformations effecting (or describing) sentential forms in this general relation are often called 'meaning-preserving' transformations; sentential structures so described are paraphrases of each other.[1]

---

[1] Whether or not such paraphrases are the 'best' paraphrases intuitively, is another question, i.e. there may be differences of opinion about whether the best paraphrase of glass^1 house^3 bird^2 would be:

(a) a bird who lives in a house made of glass

or (b) a bird who lives in a green^1 house^3

or (c) a bird who shouldn't throw stones

Without approaching such questions, we can agree that certain paraphrases rather uniformly acceptable to speakers are described as transforms in transformational grammars.

Zellig Harris first developed the notion of transformational grammars when studying the structure of discourse (28); in this work he became aware that a fundamental inadequacy of traditional grammar was the failure to make explicit clear intuitive relations among sentences of certain types (e.g., active/passive; declarative/interrogative). In order to normalize a text so that the same information was always presented in the same way he proposed a grammar that had as an integral feature the description of certain kinds of paraphrase intrinsic to the structure of languages:

> Transformations are applicable in various studies or utilizations of systems of a generally linguistic type...The chief outside uses of transformations, however, depend upon their special meaning status. Meaning is a matter of evaluation and cannot be fitted directly into the type of science that is developed in structural linguistics or in transformation theory. Still, for various purposes it may be possible to set up some practical evaluation of meaning; and with respect to most of these evaluations, transformations will have a special status. That many sentences which are transforms of each other have more or less the same meaning, except for different external grammatical status...is an immediate impression. This is not surprising, since meaning correlates closely with range of occurrence, and transformations maintain the same occurrence range. When we have transformations which are associated with meaning change, it is usually possible to attribute the meaning change to the special morphemes... in whose environment the transformation occurs.
>
> (Harris, 29)

One of the goals of transformational theory at all hands has been to explain how people perceive systematic relations among sentences. Since the transformational relationships are in large part paraphrastic, though of course there are paraphrastic relationships that are not transformational, a transformational grammar is a partial answer to the question of why people think that certain drastically different sequences of phonemes 'mean about the same thing.'

There is no explicit 'method' described in a grammar for generating or recognizing paraphrastic transforms, or anything else. However, transformational grammars do characterize such relationships in a way sufficiently explicit to support a description of the production and understanding of paraphrastic pairs. Certainly it is a fundamental empirical claim of transformational grammarians that relationships among sentential types of a transformational and paraphrastic character are in some sense 'known' or 'available' to the speaker; i.e. while the speaker may not, in general, be able to state the principles of such relationships -- in this special sense, he does not 'know' anything about grammar -- his behavior points to an implicit awareness of these relationships.

Certain theoreticians, notably Henry Hiz (33), have argued that intuitions concerning paraphrastic relationships may provide a more pervasive and reliable empirical basis for the study of syntax than intuitions concerning degrees of grammaticalness in a single sentence. In fact, the claim that people can in certain circumstances say the same thing in different ways, or recognize that two 'different' sentences mean the same thing seems to be less open to dispute than the claims concerning the ability to classify sentences according to their grammaticalness.

In 1961, W. H. Livant reported (50) on the basis of an experiment with three individuals that adult English speakers could paraphrase or 'interpret' (in effect, give transforms of) previously unheard compound-nouns regardless of semantic bizarreness, given only that they understood the individual words of the compound.

McNeill in a recent article in Science (55) studied the use of compound nouns in a different but related way. He asked engineers to derive compounds given relative clauses that might be assigned some technical meaning within engineering. Similarly, he asked them to provide the relative clause, given the compound. The subjects were to some degree able to perform both these tasks, succeeding more uniformly with the conversion to the compound nominal. However, McNeill noted that these subjects did not seem aware of the ambiguity involved in the compound, i.e. in the possibility of retrieving more than one (and meaningfully different) relative clause from a given compound.

Blumenthal (5) asked Harvard-Radcliffe students to paraphrase sentences that were instances of triple self-embedding. Most responses not only failed to be paraphrases but showed that subjects considered them to be jumbled instances of multiple conjunction.

The evidence for an effective procedure for paraphrasing, then, is also thin at this point.

Section II: Performance and Competence

It is immediately obvious that there are systematic differences between the set of sentences described in a linguistic grammar and the set of spoken and written sentences, both current and potential: Some sentences described in a grammar are too long or complex to be managed by the speaker; these are in principle unutterable. There are other expressions too foolish or farfetched ever to gain any currency anywhere, anytime; these are in practice unspeakable. On the other hand, many observed sentences are not described in a linguistic grammar. Through inattention, interference, interruption, and so forth, the speaker produces sentences that the linguist has (presumably) reason to exclude. These differences between what the grammar describes and what is said may at least in part be characterized as performance difficulties fundamentally distinguishable from the individual's competence in linguistic matters. Generative grammars, and, to a large extent, all transformational grammars, purport to describe competence; they are not direct descriptions of speech or comprehension:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory-limitations, distraction, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.
>
> (Chomsky, 1965)

Though linguistic theory therefore is intended as no more nor less than a second-order construct whose relations to actual linguistic behavior may be neither direct nor simple, it should be emphasized that the notion of competence is put forward by its proponents as a concrete psychological reality. Chomsky (1964) writes that "On the basis of a limited experience with the data of speech, each normal human [emphasis ours] has developed for himself a thorough competence in his native language." 'Competence' should be understood to refer both to the speaker's capacity for acquiring (any) natural language under reasonable conditions, and also the product of that learning, i.e. the internalized generative grammar in terms of which the speaker speaks and understands the language.

Because only performance is directly observable, the problem of giving external justification for grammars is very complicated:

> Obviously, every speaker of a language has mastered and internalized a generative grammar that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he can become aware of them or that his statements about his intuitive knowledge of the language are necessarily accurate. Any interesting generative grammar will be dealing, for the most part, with mental processes that are far beyond the level of actual or even potential consciousness; furthermore, it is quite apparent that a speaker's reports and viewpoints about his behavior and his competence may be in error. Thus a generative grammar attempts to specify what the speaker actually knows, not what he reports about his knowledge. Similarly, a theory of visual perception would attempt to account for what a person actually

> sees and the mechanisms that determine this rather than his statements about what he sees and why, though these statements may provide useful, in fact, compelling evidence for such a theory.
>
> (Chomsky, 1965)

If the analogy with visual perception is a valid one (I will not attempt to argue this question) it is clear that under no circumstances would the naive viewer come up with the statements of optical and physiological fact that indeed underlie his perception. Nevertheless, we expect that a theory of visual perception will account for what the viewer sees, once we have laid aside trivial or irrelevant interferences (he is judging by his knowledge of the known size of certain objects; he is not attending to the task; he has water in his eye). Similarly, we must ask for evidence from what the speaker says or understands, or reports concerning his speech or understanding, to support linguistic theory.

Then in order to relate language behavior to grammars, extensive psychological investigation may become necessary. At some point one must learn how to describe some of the 'false' or flawed sentences systematically, or the linguistic grammar can never be brought back to the level of data. In brief, there must be some relatively fixed relations between a linguistic grammar and language behavior if there is to be any psychologically valid explication of the notion 'linguistic competence'. Ideally, when the linguistically trivial interferences are excluded, we ought to see the grammar bare.

Because of the highly abstract and covert quality of this postulated underlying knowledge, and because of its continual interlacing in practice with other aspects of reality (i.e. those linguistically external matters affecting performance), it will turn out to be -- if it turns out to be -- singularly elusive; in fact, Chomsky says "...although one might propose various operational tests for acceptability, it is unlikely that a necessary and sufficient operational criterion might be invented for the much more abstract and far more important notion of grammaticalness."

Thus it is claimed (1) that there exists a covert knowledge of language describable as a generative grammar that accounts for the human use of language; (2) that each normal individual is in possession of this grammar, neurophysiologically; (3) that a general display of these facts will not, in all probability, be forthcoming; however that (4) there exists sufficient, in fact overwhelming, evidence from partial displays of competence both to support the theoretical account and to render further demonstrations somewhat superfluous. These claims are very strong indeed, and require from the believer a certain act of faith. The difficulty in approaching this theoretical stance is intensified when we read that, at least for some of the

Chomskian linguists, the issue has been reduced to one of definition:

> A necessary condition for something to be part of the subject matter of a linguistic theory is that each speaker be able to perform in that regard much as every other does.
>
> (Katz, 42)

We need not consider the validity of this statement for matters of performance, for there it is surely in error; if every speaker must actually perform as well as the next before we decide that a linguistic skill is being tested, then we may never discover any act that is linguistic. In fact it can be admitted in advance and without reservation that we will never encounter any task that is purely linguistic. That is not the point. It is clear from Katz' statement here (and elsewhere, cf. page 35) that his claim derives from the belief that competence is universal and equal in the normal population. If, when individual differences are found -- and they will be, as I shall show -- we are to be told, by virtue of this very finding, that they must therefore be extralinguistic, then indeed linguistic theory has become invulnerable.

Section III: An Experimental Study of the Use of Compound Nouns

In a restaurant in Philadelphia, named The Pub, there is a plaque that reads:

VOLUME FEEDING MANAGEMENT
SUCCESS FORMULA
AWARD
1963

a tribute, not only to the managers of The Pub, but also to the productivity of the noun-compounding process in English. Our admiration is only slightly damped when we discover that the search for a success formula was made by a trade-journal named Volume Feeding Management. Yet this finding makes it appear less likely that the six-word complex above arose whole-cloth as a paraphrase of:

> the award given for discovering a formula for succeeding at managing the feeding of people in volumes

Since neither phrase nor paraphrase comes trippingly off the tongue even after decipherment, we think it plausible to consider how far the use of such recursive devices is open to the individual.

Specifically, we give here a grammatical sketch of compounding processes in current English, and an experimental investigation of the knowledge and use of these processes by English-speaking adults. More generally, we are pursuing the attempt to make explicit certain relations between the language (or its grammar) and linguistic behavior. In the preceding sections, we have seen that there have been phenomenological linguists who deny any distinction between language and its users. More recently, there has developed a detailed and sophisticated model which embeds the language in the speaker. In the grammatical sketch (section A following) we provide in outline a linguistic description of compound nominals, from a transformational point of view. In the empirical work I present evidence concerning the tacit knowledge of this grammar in speakers.

A. A linguistic description of compound nominals

The term 'compound-noun' refers to certain collocations of two or more nouns pronounced as a single constituent under a particular intonation contour. The same term includes as well certain other ordered collocations of words, one or more of which may be other than nouns. The term is often extended to include certain apparently related constructions which are adjectival or adverbial in character.

1. Nominal transforms:

In English, there are many sentences containing nominals whose internal structure reflects that of independent sentences, e.g.:

That he was happy was clear to us.
The resumption of nuclear testing was greeted with hostility.
Flying planes is dangerous.
Flying planes are dangerous.

We call sequences such as those italicized above 'nominals' because, for one thing, they occupy the same positions in sentences as do simple noun phrases, e.g.:

The point was clear to us.
Macnamara was greeted with hostility.
The lion is dangerous.
Lions are dangerous.

Furthermore, the complex nominals, just like simple noun-phrases, determine other features of the constructions in which they appear, i.e. if in subject position, they determine the present-tense affix of the verb; they have selectional effects on the choice of the following verb and object, etc.

In short, it is convenient to suppose that these complex entities play a constructional role similar to that of noun-phrases in the phrase structure of the sentence. However, the sentential character of these phrases is also apparent to the speaker, e.g.

that he is happy seems related to: *He is happy*.
The resumption of nuclear testing seems related to:
  Nuclear testing was resumed.
Flying planes seems related to *Planes fly* or, alternatively,
  to *(Someone) flies planes*.

In a transformational grammar, sequences like these, having sentential character but phrasal function in the sentence in which they appear, are considered to have 'derived from' underlying sentential constructions, through formal operations that 'transform' them into nominals and embed them into other constructions. When these complex forms are treated, on the one hand, as elaborations of the phrase-structure notion noun-phrase, and on the other hand, as reflexes of the notion sentence, the description of sentence structure is kept simple, and selectional restrictions that are identical for a large number of sentential and nominal forms need not be restated, e.g.:

> Macnamara was greeted with hostility.
> Wasn't Macnamara greeted with hostility?
> Macnamara's being greeted with hostility amuses me.
> The greeting of Macnamara with hostility is understandable.
> It is treason to greet Macnamara with hostility.

Compound-nouns represent in such a grammar a further method for 'converting' or 'restructuring' a sentence as a nominal phrase, e.g.:

> Macnamara-greeting is a foolish hobby.
> Test-resumption would be folly.

2. Compounding: The particular 'process' (or relation) called compounding is a familiar one in English. In traditional terms, we speak of a compound, or of compounding, whenever two or more words or constructions are joined together in such a way that their independent status is not altogether obliterated. Thus there is said to be copulative compounding, e.g.:

> lions and tigers

in which the elements are coordinate; dependent or attributive compounding, in which one element stands in, e.g., genitive, or instrumental, relation to the other, e.g.:

the lion's foot

and descriptive compounding, in which one element qualifies or describes the other, e.g.:

houseboat
peapod

It is this last process, descriptive compounding, that is described below.

Descriptive compounding is not limited to nouns in its elements, e.g.:

black-berry
money-back-guarantee
car-wash

nor in its 'transformed' phrase-structure status, e.g.:

boot-black	(a noun-phrase)
lovelorn	(an adjective)
down-grade	(a verb)
onyx black	(an adjectival phrase)

3. Identification and pronounciation of compounds:

Compounds are centrally characterized by intonational pattern. The actual physical properties of the defining intonations are not easy to describe. Spectrographic study shows that the relations between the acoustic and perceptual facts are very complicated, in part because higher-level grammatical considerations interact with acoustic cues to determine the perception of what was said. I will speak here in terms of certain perceptions about relative pitch and stress that are more-or-less reliably associated with such constructions; but the reader should bear in mind that what the English-speaking listener reports as distinctions of relative volume (stress, or accent) and frequency (pitch) are often objectively something else, e.g. tempo changes, changes in vowel quality, elision (slurring) at the boundaries between words. Thus it is in a rather ill-defined sense that I will be discussing two major perceptual cues for compounds: (a) a pitch feature, i.e. pitch rises (↗) on the first word of the compound and falls (↘) on the last; (b) a stress feature, i.e. stress is strong or primary ($\underline{1}$) on the first word of the compound and weak or tertiary ($\underline{3}$) on the second. I will show (and it has been shown before) in the empirical work that follows that these loosely defined features in fact reliably describe the perception of compounds. For the simplest case of a two-word compound composed of monosyllabic nouns, we can write these features as follows:

[i] $\overset{\nearrow}{N}^1 \overset{\searrow}{N}^3$

for example: $\overset{\nearrow}{\text{mail}}^1 \overset{\searrow}{\text{man}}^3$

Other grammatical constructions in English often consist of the same sequences of words as compound nouns. For example, the sequence:

lady killer

may be a compound (with the meaning -- metaphorical or not -- *slayer of ladies*), or it may be a phrase (with the meaning *slayer who is a lady*). But if this latter sense is intended, the pronunciation will be characteristically different from

[i] above:

[ii] $\overset{\nearrow}{N}^2\ \overset{\searrow}{N}^1$

While the pitch feature is similar -- or perhaps the same -- the stress here is intermediate or secondary (*2*) on the first word, and primary (*1*) on the second. Such change in meaning as a function of change in stress is a pervasive feature of English, occurring within single words (e.g. *con*^{1}*sort*^{3} vs *con*^{3}*sort*^{1}) and in constructions, e.g.:

lady^{2}bug^1	(= a bug lady)
lady^{1}bug^3	(= a species of bug, both male and female)
lady^{2}bird^1	(=a bird lady)
lady^{1}bird^3	(= a First Lady)

When the stress pattern of such constructions is essentially that of the adjective-noun sequence in English (the pattern *21*), we call the collocation *phrasal*; then the term 'phrase' applies to *lady*^{2}*bug*^{1} just as it does to *old*^{2}*bug*^{1}. When the primary stress appears instead on the initial word of the sequence, e.g. *lady*^{1}*bug*^{3}, we call the sequence a *compound*.

The reader will by now have objected that the notion of fixed stress in words or constructions is something of a fiction. There are certainly occasions on which the stress patterns specified here for given meanings are not necessary, or are even wrong, e.g.

I said eye^2 brow^1, not eye^2 lash^1.

There are, then, overriding considerations of sentential stress (ordinarily called 'emphatic' or 'contrastive' stress patterns) that cause the usual, or 'colorless' or 'normal', stress patterns to be superceded and effectively submerged. As I will show by responses of subjects, the colorless stress pattern is ordinarily assumed by the speaker in the absence of contrastive contest.

Another difficulty in specifying the stress of compounds arises because there is a very definite and very small number of distinctions of stress and pitch that can be maintained reliably by listener and speaker. Then as soon as the compound becomes longer (has polysyllabic elements or has more than two elements) these features become confusing and distinctions wash out. There are three, just possibly four, distinctions of stress that are produced and perceived reliably; there is rising, steady, and falling pitch; there are in addition a few perceptible changes in tempo and vowel quality. The objective details of this intonation system are not altogether well known. What is clear is that as the compound grows longer there are insufficient distinctions available in the language for keeping some sequences apart. For example, while we can distinguish between the two cases of lady bug, we cannot so easily tell whether a lady^1 bug^3 house^2 is a house for the (species) lady bug, or a house for female bugs. Occasionally, even in this simple case, a speaker may report the paraphrase *asylum for females*, suggesting that he heard *lady*^{2} *bug*^{1} *house*^{3}. These difficulties for the hearer rapidly become insurmountable; it is impossible, without further context, to know what an *old lady bug house keeper* does for a living.

4. Grammar of compound nominals: In modern grammars, the kinds of phrases and compounds of which we have been speaking are ordinarily described by what are called transformational rules. These are, very roughly, statements describing relations among sentential structures. In generative grammars, more complex structures are derived from simpler sources, and further 'derived structure' is assigned to the complex expression. Presumably the speaker is in some tacit sense aware both of the immediate ("surface") structure of the sentences he hears or says and also aware of the ("deep structure") sources of these sentences, i.e. the simpler structures from which they are derived, and the mode of that derivation. For example, the listener who hears the sentence

John is approached by spies.

is aware that *John* is the subject of that sentence, in the sense that the singularity of the word *John* determines that the following word must be *is* and not *are*. At the same time, he knows that *spies* is in some deeper sense the subject of the sentence, i.e. that the Passive is related in a very definite way to the Active structure underlying:

Spies approach John.

The fragment of a transformational grammar describing compound nominals thus attempts to explain the sense in which

a wood^1 box^3 (a compound-noun)

is palpably akin to such apparently different constructions as:

a box for wood (a prepositional phrase)
and a box that holds wood (a relative clause)

Further, a complete grammar would seek to explain some rather more distant relatives such as:

A box holds wood.

and even:

Does a box hold wood?
Is wood what a box holds?
A box does not hold wood.

The network relating all these forms does service, of course, for an unbounded number of sequences differing morphologically, but not structurally, from this example, e.g. brick house; teakwood box, thus -- if successful -- describing at once the wealth of possible utterances of the language and the organizational principles explaining their rather uniform comprehensibility.

a. Compound Nouns[1]: We have noted that the immediate paraphrastic kin of compound nouns seem to be relative clauses and prepositional phrases. The relationship can be roughly stated as follows:

$$[\text{iii}]\ N_1 \left\{\begin{matrix} P \\ \begin{Bmatrix}\text{that}\\ \text{wh-}\end{Bmatrix}\ \text{(be)}\ \ V\ \ \text{(+en by)} \end{matrix}\right\} N_2\quad \text{(+plural)} \rightarrow N_2^{\overset{\nearrow}{1}}\ N_1^{\overset{\searrow}{3}}\quad [=N]$$

where N = noun; V = verb; P = preposition; + indicates an affix; braces indicate a choice; parentheses indicate optional elements; → is the relation said to hold between the two sides of the formula. The bracketed note following the formula indicates the structure assigned to the right-hand (output, derived) expression.

For example:

a gear that drives a wheel - a wheel gear
a gear for wheels - a wheel gear

The source or prior structure for compound-nouns is thus said to be a noun-phrase containing a relative (that), or a preposition. The structure of outputs (the right-hand form) is again a nominal structure, a unit: N. This will mean that the rules are reapplicable, using their own outputs in further inputs. For example, if wheel gear is an output of [iii], that output is a noun, so it may appear in the left-hand of [iii], e.g.:

---

[1] The grammatical description given here is neither complete nor wholly original. Much is owed to R. B. Lees (45) who is nevertheless not responsible for errors here. The material is presented to provide context for the experimental work and is in no sense a "definitive" grammar.

a wheel gear for trains	- a train wheel gear
a train that uses a wheel gear	- a wheel gear train

Thus a single formula can describe all the (infinitely many, arbitrarily complex) expressions of this type, at the same time providing a statement of the syntactic properties of all its outputs, i.e. predicting the constraints on its relative occurrence within sentences.

Notice that if N_2 is plural, the plural affix is lost in the compound form. This will mean, in the general case, that nonfinal elements of compound nouns have -- like adjectives -- no inflexional affixes.[1] The mechanism of compounding, essentially, is inversion of the two nouns, marked by the addition of a characteristic intonation pattern.

---

[1] Because the plural affix is in part homophonous with the genitive affix, certain examples (e.g. calves liver) appear to be exceptions. However, particularly in compounds with genitive sources (which will not be discussed here), there are unambiguous exceptions, such as women's club. There may be isolated real exceptions when [iii] is in fact the syntactic relation involved, e.g. purple people eaters. At any rate, such exceptions seem to occur only in case the noun involved does not take the morph +s as plural, but has instead an 'irregular' plural. Thus purple person eaters (in either reading) is permissible, if stodgy, while purple persons eaters is ungrammatical.

It is clear from formula [iii], that the 'conditions' or prior structures associated with compounds are many and varied. Then very different nominal expressions are matched by the same compound, e.g. in principle a *horse cart* might be:

a cart that is shaped like a horse (as in *box car*)
a cart that is pulled by a horse (as in *dog sled*)
a cart that a horse rides in (as in *hay wagon*)
a cart for a horse (as in *dog house*)
a cart that is as big as a horse (as in *horse radish*)

Further, there may be 'set' or 'frozen' meanings of the same phrase that are only metaphorically -- if at all -- related to the independent meanings of the element words of the compound as in, e.g. *horse play* (boisterous play), *horse sense* (common sense), *cartwheel* (an acrobatic trick), *Charley horse* (stiffness from muscular strain). We shall have more to say concerning both the ambiguity of compounds and the character of set phrases after further "rules" for regular compounding are discussed.

b. compound nominals with verbal components: Further compounds are formed from the same source structures by essentially the same mechanisms (inversion and addition of the characteristic intonation pattern) by deletion of a nominal rather than the verbal or prepositional elements:

$$[\mathrm{iv}]\quad N_1 \begin{Bmatrix}\text{that}\\ \text{wh-}\end{Bmatrix} \text{V+tense} \quad N_2 \rightarrow \overset{\nearrow}{N}{}_2^1\,\overset{\searrow}{V}{}^3 \quad [=N]$$

for example:

a gear that drives a wheel - a wheel drive

$$[\mathrm{v}] \quad N_1 \left\{\begin{matrix}\text{that}\\ \text{wh-}\end{matrix}\right\} \text{V+tense} \quad N_2 \rightarrow \overset{\nearrow}{V^1}\,\overset{\searrow}{N_1^3} \quad [=N]$$

for example:

a gear that drives (a wheel) - a drive gear

c. compounds with copulative verbs: Certain other compounds seem to be associated in a very similar way with relative clauses containing the copula be:

$$[\mathrm{vi}] \quad N_1 \left\{\begin{matrix}\text{that}\\ \text{wh-}\end{matrix}\right\} \underline{\text{be}}\text{+tense} \quad N_2 \rightarrow \overset{\nearrow}{N_2^1}\,\overset{\searrow}{N_1^3} \quad [=N]$$

for example:

a friend who is a girl - a girl friend

d. naming: Other compounds are paraphrastically related to clauses that provide names, e.g. Leyden jar, Main Street; and to species names, e.g. elm tree, hound dog, blueberry pie. These are perhaps most easily described as an extension of [vi], which will be restated below.

e. compounds with adjectival elements: Noun phrases with adjectival modifiers can be described through their relationship to predicative phrases, i.e.:

$$[\mathrm{vii}] \quad N \left\{\begin{matrix}\text{that}\\ \text{wh-}\end{matrix}\right\} \text{be+tense} \quad A \rightarrow A^2\, N^1 \quad [=N]$$

for example:

a man who is stout - a stout man

Notice that the stress is phrasal, not compound. However, there are very many cases of adjective-noun collocations which have instead the compound stress patterns, e.g. blackbird, White House. These constructions have obvious relations to the naming processes sketched in section d above. The meanings of these phrases are frozen in rather specific ways. Thus, for example, while

a pure^2 blue^2 bird^1

must certainly be blue from tip to tail,

a pure^2 blue^1 bird^3

may (and does) have a distinctly reddish breast. Certainly the blue^1 bird^3 was so named because it is a bird that is blue, but the expression blue^1 bird^3 does not refer to all birds who are blue, but only to a particular species. No rule of grammar can predict whether an adjective-noun pair is to be pronounced as a compound or as a phrase. In effect, this is what is said by lexicographers when they list and define the compound blue^1 bird^3, but do not list the phrase blue^2 bird^1. The latter can be understood by reference to its element words and rule [vii]; the special meaning of the former must be given lexically. The difficulty in this situation is the openness, or productivity, of the naming operation. Where elements, or even constructions, are finite in number and relatively fixed in membership, simple listing is more or less sufficient descriptively. However, recursively defined constructions cannot in principle or in practice be listed exhaustively. Hence, for Language and for speaker, it is assumed that some rule such as [vii] is involved. In this instance, what are not described are the semantic conditions that may give rise to the compounding stress pattern. Lees (op. cit.), in a very revealing discussion of these matters, points out that there are naming operations that maintain phrasal intonation without being in any structural or semantic way obviously different from those taking compound intonation, e.g. *Near East*, *Tiny Tim*, *common cold*, *Big Ben*. It suffices here to provide the general syntactic picture by generalizing rule [vi]:

$$[\text{viii}]\quad N_1 \begin{Bmatrix}\text{that}\\ \text{wh-}\end{Bmatrix} \text{ be+tense } \begin{Bmatrix}\text{named}\\ \text{like}\end{Bmatrix} \begin{Bmatrix}N_2\\ A\end{Bmatrix} \rightarrow \begin{Bmatrix}\overset{\nearrow}{N}{}_2^1\ \overset{\searrow}{N}{}_1^3\\ A\end{Bmatrix} \quad [=N]$$

f. frozen compounds and the 'appropriate' verb: In section *a* above, I listed a number of theoretically possible meanings for the compound *horse cart*. However, a horse cart is just a cart that is drawn by a horse. We may know this either conventionally -- because the compound has become a lexical item with a set or frozen meaning -- or because of certain semantical relations between the elements of the compound that lead to only one or a very few 'sensible' or 'likely' things that it could mean.

1) set compounds: We know that fly paper is paper for catching flies, not paper that flies write on or paper that flies. This is in no sense given in the syntax of the compound. A particular meaning has been 'frozen onto' the compound, even though its underlying structure is still apparent. We know why fly paper is called fly paper, but it will contine to be called fly paper even if we catch spiders in it; further, we will probably avoid calling paper that flies write on fly paper. Such set compounds must be entered into a grammar as lexical items, and they may be expected to be handled psychologically in ways quite different from compound constructions. Work with children (Livant ; Berko) shows that frozen compounds are, in general, perceived initially as simple lexical items, and their constructional features are only 'discovered' later in development, and very possibly by reasoning processes that are largely extralinguistic in character. (My children tell me that a black board is called a black board because you write on it.)

2) appropriate verbs: Zellig Harris (3c) has pointed out that there are cooccurrence relations in the language that serve to disambiguate compounds formed on the basis of rule [iii]. Your mail man is the man who brings your mail; the man who burns your mail is not your mail man,

unless, perhaps, you are with the CIA. The man who burns your leaves in the autumn might conceivably be called your _leaf man_, but the vandal who burns your leaves in the spring is not your leaf man. Certain verbs are very generally associated with compounds (e.g. _bring_, _yield_, _look like_), and certain other expressions will always 'do', in a general way, although they are not very revealing (e.g. '_in some way characteristic of_'). The apparent generality of some of these associations has led to some attempts to choose some such verb or phrase as a specific to the compounding process, i.e. to substitute for the _V_ of formula [iii] a specific item. This enterprise does not succeed very well, yet it fails to fail altogether; this state of affairs is awkward. I refrain here from any attempt to characterize the notion of appropriate verbs syntactically; we shall see, however, that it has a great deal of vitality as a psychological concept.

g. exocentric and endocentric constructions:

A _dog house_ is a house, of a kind. It is just that kind of house that dogs live in. The cooccurrence restraints (or privilege of occurrence) for _dog house_ overlap considerably with the restraints on the cooccurrence of _house_, though there are some gross differences. On the other hand, the cooccurrence range of _dog_ is not so nearly like the cooccurrence range of _dog house_ as is _house_. In this sense, _house_ is the 'head' of the phrase, *dog* a limiting modifier.[1] Compounds whose elements stand in this modifier-head relationship are called *endocentric*.[2]

On the other hand, the cooccurrence range of *cut throat* is close neither to that of *cut* nor that of *throat*. It is close to that of *hoodlum* or *thug* or *hood*. This kind of compound is called *exocentric*. There are compounds on most of the models we have described that are exocentric, e.g. *mouthpiece* (a lawyer), *deadbeat* (a sponger). Exocentric constructions are by definition frozen, because their meaning cannot be reconstructed by reference to the lexical items they contain and internal features of the construction. That this is so is a matter of degree. An ingenious speaker who had never heard the words cutthroat or deadbeat might conceivably figure out the senses in which these words are intended in speech. But these meanings are not 'pregiven' by reference to a regular paraphrase. It is because the constructional features that (at some time or other) gave rise to the compound are retrospectively clear (i.e., if we know that a cutthroat is a hood, it is obvious that a cutthroat was called a cutthroat because he cut people's throats) that grammarians are reluctant to demote these constructions to the status of mere words. It is a psychological question, open again for each such presumed compound and open again for each speaker, whether the compound has any perceived underlying constructional stature.

---

[1] There are additional reasons to call the second element, rather than the first, in such constructions the head. The second element will determine, for example, whether the compound requires an article (*the*, *a*), e.g.:

We prefer the rice package. but not *We prefer rice package*.

We prefer the package rice. and also *We prefer package rice*.

This distinction can be accounted for by assuming (1) that *rice* is the kind of noun that occurs freely in the singular without an article, e.g. *We prefer rice.* and also *We prefer the rice.*, while *package* is not; and (2) that the second member of the compound is its head.

[2] In describing the distinction between exocentric and endocentric compounds I am following Lees (op. cit.).

All exocentric constructions are treated here as mere words. This decision is a strategic one, thus in part arbitrary. The problem of treating these items as constructions is that we are then faced with an immense and unmanageable continuum. Consider, for example, the problem of deciding which of the following compounds with dog are real constructions (only exocentric) and which are frozen compounds (i.e., lexical entries): dog ape (a baboon), dog bane (genus Apocynum), dog cart (a cart drawn by a dog), dog days (the sultry part of the summer), dog fight (a melee of airplanes), dog fish (any of several small sharks), dog paddle (a swimming stroke similar to that used by dogs), dog rose (Rosa canina), Dog Star (Sirius), dog tooth (the tooth of a dog), dog trot (a gentle trot, a dance), dog wood (Cornus), dog watch (a watch of two hours on shipboard). I consider a compound a compound only in case (1) the transformational paraphrastic relations for the item are appropriate and (2) the usual alternative interpretations are appropriate and available. Thus dog cart is a compound only in case it may be a cart in which a dog rides as well as a cart which a dog draws.

h. moribund processes: In the preceding sections, I have described what I believe to be the major productive processes of compounding current in American English. There are, in addition, certain classes of apparent compounds constructed according to models that are no longer in active use. Major examples are verb-object compounds (such as cut-throat, pickpocket, shunpike) without the inversion of order common to all the active processes described. I will contend that English speakers no longer make use of this model for the creation of new nominals in the language.

i. comparative modification: A further construction is described here, though it is perhaps unrelated to the constructions under consideration. Its outputs are, however, often strings identical to compounds and nominal and adjectival phrases simply because of the extensive lexical ambiguity in English between common nouns and adjectives. Such constructions will show up in the empirical work sufficiently often so that we ought to have a name for them:

[ix] as A as N → $\overset{\nearrow}{N}^2\ \overset{\searrow}{A}^1$ [=A]

for example:

as cold as ice - ice cold
as red as a rose - rose red

Notice that these adjectival phrases are subject to nominal interpretation (a Communist rose).

j. parts of speech. Lexical assignments are represented in a grammar as, e.g.:

[x] N → bird; house; man;...
[xi] A → glad; wry; sick; ...
[xii] V → eat; sing; speak;...

As we have seen, there are ways of adding members to lists like [x], e.g. through rule [iii]. A recursive rule is used rather than simply adding, e.g. *horse cart*, to [x] because, as [iii] implies, there are arbitrarily many complex members of [x].

Some simple words are cross-classified, i.e. they appear on more than one lexical list. For example, *dry* is used both as adjective and as verb. There are also cases of words that appear on one list and appear again with a fixed addition (affix) on another list. For example, many verbs act as nouns in sentences when they have the suffix *-tion*. Since, to a limited extent, it is possible to add *-tion* to words and constructions that never had them before, it is not only convenient but necessary to describe this feature of the language by rule; otherwise an exhaustive lexicon cannot be provided.

There is a case intermediate between cross-classification and change of class-membership by affixation: often a word is used in more than one way without the addition of an affix; yet one use seems intuitively 'basic', the other less so. For example, *yellow* is clearly an adjective. Yet it may be used as a nominal (*Yellow is my favorite color*) and even as a verb (*The parchments yellow*). If the grammar that is correct is just that grammar that describes the speaker's intuition about the language, it would be convenient to mirror the feeling of 'basic classification' as a simple list-entry, and to derive other classifications by rule (i.e. by assuming that *yellow* as a noun or verb was a special case of affixation). There is no formal justification for such a move, but we will see that there is ample evidence of intuitive knowledge of basic classification.

B. An empirical study of the use of compound-nouns

Three general questions are raised in these experiments:

(1) In what ways is a linguistic grammar describing the capacities of individuals? Although the informal evidence for some projective device in individuals is overwhelming, a private grammar may be more limited than that describing the Ideal Grammarian. Certainly the view of man as algebraist is rather novel in psychological circles (see, however, recent formulations, e.g. Miller et al. 1963; Neisser, forthcoming).

(2) If people 'have' such grammars, what linguistic activities does this make possible for them? Can they make decisions about grammaticalness? Can they paraphrase? We know that a grammar is a description of something now called 'competence', but competence at what?

(3) How uniform are people in their organization of the language and in their linguistic activities? If there are no significant differences in adult linguistic behavior (as, e.g., Chomsky, Putnam, Lenneberg, and Katz suggest), then we can fix the relations between the grammar and speech in essentially simple ways.

It has seemed reasonable to many linguists to endow each person with all the wealth and richness of the language, regardless of the fact that this undercuts any broadening of the aims of linguistic theory: it would seem on the face of it that the study of style, fluency, and articulateness would be related at least in part to differing approaches and capacities in matters of grammar. If all speakers must be alike in some aspect of language use for that aspect to be part of linguistic theory at all, then linguistic theory cannot be counted on to provide insight into features of language use that presumably concern us. This seems a highly arbitrary and unjustified limitation to place upon linguistics, one that would rob it of texture and descriptive capacity.

The particular aspect of language use I study here is that of compound-nouns, and certain related structures in English. Compound-nouns offer a very convenient aspect of English for study. Their complexity and internal structure can be varied considerably without adding to the phrase either in length or in vocabulary. Compounds are often ambiguous as strings, overlapping with various other nominal and adjectival structures, partially because of the lexical ambiguity, in English, of most of the short words in the open lexical classes. Compounding at a considerable degree of complexity is a frequent and familiar productive activity in English, one that shows up at a relatively early stage in linguistic development, and one that is used without restraint even in the most rudimentary discourse:

> I like to box.
> How I like to box!
> So, every day,
> I box a Gox.
>
> In yellow socks
> I box my Gox
> I box in yellow
> Gox box socks.
>
> (Seuss, 1960)

With these materials, I have studied the way individuals paraphrase under experimentally controlled conditions. In earlier discussion (cf. Section I), I suggested why paraphrasing might legitimately be considered a feature of behavior relevant to linguistic competence. It goes without saying that the proponents of the notion 'competence' may disagree with this estimate, and they are certainly free to do so, although they have themselves rested much of the claim for grammatical intuition on the presumptive ability to paraphrase. In my opinion, the ability to perceive sentential relationships of a formally statable kind -- and that is what paraphrastic transforms are -- would be a basic prediction of the theory of an internalized generative grammar, once it is granted that this hypothetical competence makes the individual competent to do anything at all with language.

In The Structure of a Semantic Theory (43), Katz and Fodor comment concerning the understanding of complex nominals:

> It is only because there is a systematic way to understand the meaning of such constructions in terms of the meanings of their parts that the infinite stock of strings produced by the grammatical mechanism for creating new modifier-head constructions can be employed by speakers to refer to familiar objects.

The assumption that will concern us in this is that speakers are presumed to understand a compound (excluding moribund and frozen forms) if they understand the separate words comprising it and, of course, if they understand the structure of the compound.

I have mentioned briefly an experiment by W. H. Livant (50, op. cit.) with three individuals that purported to show that English speakers could paraphrase or 'interpret' previously unheard compounds regardless of semantic bizarreness. A paraphrasing task of the kind Livant suggests would provide the first kind of evidence that ought to be looked for in support of the theoretical remarks of Katz and Fodor.

Livant's stimuli were three-word sequences all containing the same words (black; bird; and house) in all six orders and with two stress patterns common to compounds (132, 213), for a total of twelve stimuli. He reports that his subjects were uniformly successful in providing paraphrases. The implication of this result is that speakers are able to generate at will sentential structures related (in the grammar) in very specific ways to compound nouns and certain other nominal and adjectival phrases.

It is worth submitting this issue to a more general and rigorous test. Uniform success in this task would give credence to the rather common view among transformationalists that there is little individual difference in adult linguistic ability. A finding of asymptotic similarity in adult linguistic competence would strongly support the view of language as 'instinct'. Such results would give detail and texture to the notion of grammatical competence as a psychological phenomenon. It is plausible that the individual 'in possession of' a transformational grammar might recognize meaningful and grammatical relationships among sentences with closely related morphology and deep structure. An empirical display of the fact that people can generate paraphrases at will without contextual suuport would be a result of great strength concerning the breadth and conscious availability of grammar. That adults could uniformly provide a particular kind of paraphrase under rather unnatural and difficult circumstances would be startling.

In this study, as in any laboratory situation, we can expect the results to be confounded with all sorts of interferences that can be called, for convenience, 'performance errors' unrelated to what we set out to study. Although we cannot exclude these factors, we can hope to produce a situation in which such extraneous effects can be extricated from the phenomenon under investigation.

1. Pilot work: In the course of studying the use of compound-nouns by children, I attempted to replicate Livant's results with adults. Livant does not report the conditions under which he presented the stimuli to his three subjects, but I assumed that the only realistic test would be an oral presentation (because some of the grammatical cues for these constructions are not reflected in English orthography). I was not able to reproduce his results: where his subjects had uniformly succeeded in providing acceptable paraphrases for the compounds, none of my twelve subjects performed perfectly. Table I summarizes the results of this work.

Table I shows that where the compound can be derived using only rules [iii] and [vii] or [viii] subjects give a correct response over 80% of the time. Where the derivation involves any other rule ([iv] or [ix]), subjects are correct only about 29% of the time.

Notice that errors are associated here most strongly with the requirement to use rule [iv]. Is rule [iv] 'difficult'? Not necessarily: it is only that neither bird, black, nor house is a verb. It is possible to view black as a verb as a verb (i.e. to add *black* as a member of rule [xii]), but this requires at a minimum an extra computation, and at a maximum an exercise of ingenuity.

The simplest description of these facts is: where *black* can be viewed as Adjective, the responses are correct most (90%) of the time; where *black* is viewed as Noun (color-name), responses are correct less often (62%), and where *black* is viewed as Verb, responses are rarely (12%) correct.

Each subject in this experiment did best when *black* was used as an adjectival or nominal modifier; there was no exception. Each subject did worst when *black* could be interpreted as a verb. Those subjects who found a legitimate paraphrase when *black* was used as a simple noun were the only subjects who found a legitimate paraphrase for *any* instance of the verbal use of *black*. All subjects who had a college education found a legitimate paraphrase when *black* was used as a simple noun; no subjects without a college education found a legitimate paraphrase when *black* was used as a simple noun.

My first supposition was that less sophisticated subjects might have had trouble understanding or accepting this strange task. Some of the responses called for are very bizarre (*a bird who blackens houses*). All understood the task to at least this extent: they gave responses that were noun-phrases containing a relative clause. But not all matched

Stimulus	Sample correct response	Subjects who gave correct response		Rule describing:	
		#	%	intern. compnd.	extern. compnd.
1. bird^2 black^1 house^3	a black house for birds	10	83.3	viii	iii
2. black^2 house^1 bird^3	a pet bird that is black	12	100	iii	vii
3. house^1 black^3 bird^2	a bird who blackens houses a bird as black as a house	2	16.7	iv ix	viii viii
4. bird^2 house^1 black^3	a bird who blackens houses	0	0	iv	viii
5. house^2 black^1 bird^3	the black bird that lives in the house	12	100	viii	iii
6. bird^1 house^3 black^2	as black as a bird house stuff for blackening houses	8	66.6	iii iii	ix iv
7. bird^1 black^3 house^2	a house which is black as a bird a house where those who blacken birds live	2	16.7	ix iv	viii iii
8. black^1 house^3 bird^2	a bird who lives in a blackhouse	10	83.3	viii	iii
9. house^2 bird^1 black^3	stuff for blackening birds that is kept in the house; one who blackens birds who lives in the house	2	16.7	iv	iii
10. house^1 bird^3 black^2	that black that is characteristic of certain housebirds stuff for blackening the bird that lives in the house	7	58.5	iii iv	ix iii
11. black^2 bird^1 house^3	a birdhouse painted black	11	96.6	iii	vii
12. black^1 bird^3 house^2	a house for birds who are black; a house for blackbirds	10	83.3	viii	iii

Table I

Results of Replicating Livant's Study with 12 Subjects the 'right' nounphrase with the stimuli.

As a first check on the possibility that some subjects simply lack the ingenuity to come up with an acceptable response, I retested two subjects; as follows. I told these subjects that there were 'right' and 'wrong' answers. I repeated the stimuli, giving the subject the right answer, and asking him to compare this with (his own) wrong answer. Though they were willing to agree that my answers were 'better', on a retest an hour later they made the same errors.

2. Experiment 1: the free generation of paraphrases:

I was now persuaded that there might be a real difficulty, for some subjects, in understanding complex compound nouns without strong contextual support. At the same time, I had found no subject who had trouble with the task of paraphrasing itself: unless the stimulus was complex in some very special way, each subject had been able to provide legitimate paraphrases.

I therefore decided to test (1) the effect of educational level (hence, hopefully, intelligence) on performance in this kind of task, (2) the effect of structural complexity of the stimulus and, (3) the effect of semantic interference with a structurally determined response, i.e. the tendency -- in the pilot study -- to say bird^1 house^3 as part of almost any wrong response.

a. stimuli: In the pilot-study, it seemed that the subjects did best when *black* could be interpreted as a prenominal adjective.

Yet the stress and position of the word *black* in these compounds does not always permit this simple interpretation. In that case, either the subject is going to come up with a more complicated interpretation or he must give what I will call a 'wrong' response.

To test this general hypothesis, we chose members of various lexical categories permissible in compounds: *noun*, *verb*, *adjective*. Some words chosen belonged to more than one category, so that there was a choice of interpretations. For some of these, we could decide intuitively on a primary (basic) classification and a secondary (derived) classification. Presumably, the correct answer is harder to come by if the structure of the compound requires the use of one of these derived category memberships.

We chose twelve words to vary with *bird* and *house* (these, in the spoken language, unambiguous nouns). Each word is simple (according to Thorndike and Lorge (76)), monosyllabic (to avoid internal stress complications), and has no inconvenient morphophonemic interactions with *bird* and *house*. All stimuli contain the words *bird* and *house*, and one of the 'test' words. All six orders of three words and two stress patterns

(132, 213) occur. Then there are 144 stimuli. The stimulus list appears as Appendix A. All 144 were presented to each subject. Table II gives the twelve test words, and their presumed linguistic classifications. Here and there, there may be some disagreement with the classification given in Table II. The reader may, for example, think that the use of glass as noun is not as 'basic' as its use as adjective. Further, it is clear that glass and black, both classified here as adjectives, differ in some major respects syntactically.[1] There is no need to resolve these issues in advance of the experiment.

Of the 144 stimuli, there are grammatical interpretations for 132; the 12 stimuli containing the word eat are ungrammatical, or at best awkward.

We recorded the stimuli on tape so that all subjects received the identical acoustic information. The stimulus-tape was pretested to assure that the relevant acoustic information (stress, clarity, etc.) was there to be heard.

---

[1] You can say a very black house but not a very glass house, a blacker house but not a glasser house or a more glass house.

Test word	Basic classifications	Derived classifications
black	adjective	noun, verb
stone	adjective, noun	verb
glass	adjective, noun	
dry	adjective, verb	
thin	adjective	verb
bright	adjective	
boot	noun	verb
foot	noun	
shut	verb, past participle	adjective
wash	verb	noun
kill	verb	noun
eat	verb	

Table II

Lexical classification of test words.

To this end we taught two of our staff (one clerical, one professional) how to indicate stress, and had them transcribe the stimulus tape. There was 95% agreement. The 5% of stimuli that were not described indentically by these judges were recorded and retested until complete reliability of transcription was achieved.[1]

b. order of stimuli and presentation: Stimuli were randomized by blocks of twelve according to test-word and also according to 'structural type' (i.e. stress and order of words).

Subjects in each experimental group (see below) were divided at random into three order-groups. For subjects in order-group λ, the tape was started at the first taped stimulus; for those in group $\underline{\beta}$, the tape was started at the 37th stimulus; for those in group γ at the 49th stimulus.

---

[1] There is more than one intonation contour associated with each of the stress patterns used here; probably they differ in tempo. We studied some of these contours to see whether subjects could distinguish between them. We found that subjects could distinguish between the pattern intended for *pea^1 soup^3 green*^{2} (*as green as peasoup*) and *pea^1 soup^3 green*^{2} (*the vegetable you put in peasoup*). Reliability in making this distinction was well over 80% for three respondents. The stress-pitch patterns we chose for study were those of *peasoup green* (*the vegetable*).

Stimulus 1 was unfamiliar and semantically odd but grammatically possible (foot^1 house^3 bird^2), Stimulus 37 was partially familiar and semantically transparent (stone^2 bird^1 house^3), and Stimulus 49 was unfamiliar, bizarre, and ungrammatical (eat^1 bird^3 house^2). Varying the order of presentation allowed us to check whether a subject might be affected all through the test by particular features of the first stimulus he faced. Since the order of presentation was changed by starting the same tape in different places -- not by using other recordings -- all subjects nevertheless finally heard physically identical stimuli.

c. subjects: All subjects were female, essentially monolingual English speakers; most were mid-Atlantic and mid-Western in background, but one was Canadian. All were between the ages of 19 and 36.

Group A:	Seven graduate students in various fields.
Group B:	Seven undergraduates and college graduates who had no intention of doing graduate work.
Group C:	Eleven high-school graduates who had no intention of going to college.(secretaries).

All subjects were paid at the same rate for participation in the study. All were told that they were 'participating in an experiment about the way people use English'. Four subjects from group C were later discarded because of possible failure to understand the directions. The criteria for discarding subjects are discussed later.

d. instructions: The experimenter read the instructions to each subject. If the subject seemed confused, the experimenter was to repeat and amplify -- but not add other kinds of information. In fact, the subjects seemed to understand the directions as they were given:

> This is an experiment on English compound nouns. Examples of English compound nouns are *milkman* and *money back guarantee*.
>
> I will play some compound nouns on this tape recorder. When you hear one of these compounds, I would like you to give me a phrase that means about the same thing. For example, if I say *milkman*, you would say *a man who brings the milk*, or something like that. If I say *money back guarantee*, you would say, *a guarantee that you'll get your money back*. I do not want you to just use the same words in another sentence. For example, the sentence *Milk falls on the man* contains the words *milk* and *man*, but it does not mean the same thing as *milkman*. *A man who brings the milk* *does* mean about the same thing as *milkman*. *A guarantee that you'll get your money back* *does* mean about the same thing as *money back guarantee*.
>
> All the compounds you will hear will be three words long. To make sure you have no problems about hearing them, here is a list of the words that will appear in various combinations of three. [subject is here given a typed list of all the words]. No other words will ever be used.
>
> Some of the compounds you hear may be familiar to you, and you may know what they mean. Others may never even have been said before by anyone. Try to think what they might mean if they were ever used. Sometimes this may be something that strikes you as silly or even impossible. For example, I am sure you have never heard the words *duck*^{1} *food*^{3} *man*^{2}. Try to think what it might mean if it were used. I guess it would be *a man who makes duck food*, or *a man who brings food for ducks*. Is this clear?
>
> You will hear about 150 of these. Each time you will give me a sentence or phrase that means about the same thing as the compound you hear. If you want me to repeat one of them for any reason, just ask. Take as long as you like to answer.

Sometimes people ask me whether they should pay attention to the way we say the compound. The answer is "yes", because the way we say it may affect the meaning. If you become tired, tell me, and we will take a break. We will take a break after each 50 anyway.

Remember: I want you to give me a phrase or sentence that means about the same thing as the compound noun you will hear. I will remind you of this automatically after each 12, so don't think that my reminding you indicates that you have been doing anything wrong.

After the instructions are given, the experimenter plays the first stimulus. Now the subject invariably asks a question of this general form: "Now? Is this what? Now I'm supposed to say what that means?" The experimenter says "Yes". Now the subject gives a response, and then usually asks: "Is that right? Is that what you want me to do?"

The experimenter answers (regardless of whether the subject got the right answer or the wrong answer):

Yes, if that is a sentence or phrase that means about the same thing to you. I'm supposed to keep a straight-face here, and not show whether or not I agree with your answer, so don't think your answers are wrong because I don't nod or say 'yes' or anything like that.

The following portion of the instructions is repeated after every twelve stimuli, unless or until the subject gets annoyed and says he understands perfectly well and is bored with the repetition of instructions:

I am now repeating part of the instructions, as I told you I would. Remember: I want you to give me a phrase or sentence that means about the same thing as the compound noun you will hear.

e. procedure: After reading the instructions, the experimenter played the stimuli one-by-one, and the entire experimental session was recorded on another tape-recorder. After every 50 stimuli (or when the subject requested) a 15 minute break was taken.

Subjects sometimes repeated the stimulus aloud before or after giving a response. Occasionally the repetition was inaccurate, i.e. the subject misreported the stress or the order or even the word itself (foot was misheard as boot). In these instances, the experimenter said "Let me play that one for you again," and the subject was encouraged to respond again.

After the experiment was done, the experimenter asked the subject his age and educational background. These questions were delayed until after testing because subjects asked about their educational background may assume they are taking an Intelligence Test (perhaps they were).

f. scoring:

1) transcription: Subjects' responses were transcribed onto form-sheets from the tapes by two independent judges. There were a very few instances in which one or the other transcriber had trouble understanding what a subject had said. Otherwise, the results of transcription were reliable. (Greater than 98%).

2) correct vs. incorrect response: A linguist and a graduate student in English Literature were trained to score the responses, on the basis of the grammatical descriptions provided in Section and above and on the basis of their intuitions concerning the correctness of the response. Both judges scored the responses for three subjects so that reliability of scoring could be checked. There was greater than 95% reliability between the two independent scorings made by these judges. Item by item scoring was checked to see whether a response scored as correct for one subject was scored the same way for other subjects. There was a slight (not statistically significant) 'halo effect', i.e. a tendency to score 'good' subjects leniently, possibly because their responses were more fluent and literate. Inconsistencies of this kind were corrected. It is curious that the stimuli that we call ungrammatical (those containing eat) were not a source of unreliability in scoring: in these instances, it is difficult to provide any external description of what is called a correct or an incorrect response; yet judgments were reliable. The validity of such judgments is a separate question.

Subjects were considered to have made the correct response unless it was definite that they had made the wrong response.

In cases of ambiguity, the response was scored correct where one choice among the ambiguities was correct. For example, if *bird^2 black^1 house^3* got the response: *a house for a bird that is black*, the answer was scored as correct. Little attempt was made during the experimental sessions to ask subjects to clarify such responses, because pilot work showed that any 'probing' tended to confuse and mislead the subject.

The particular verb or preposition chosen by the subject when giving responses related to rule [iii] (where the verb or preposition is not present in the compound) was not considered of particular relevance in deciding whether a response was correct or incorrect. For example, in response to *black^1 house^3 bird^2*, a subject might say:

> a bird who lives in a black house
> a bird from a black house
> a bird for black houses
> birds who hang around black houses
> birds who build their nests in black houses
> a bird of black houses

All of these responses are equivalently described by rule [iii], and are considered the identical response. Responses are often expanded, i.e. the internal compound is also paraphrased, e.g.:

> a bird that lives in a house that is black
> a bird that lives in a black^1 house^3, and it is called
>   a black^1 house^3 because it is a black^2 house^1

Expansion of the internal compound was not required in scoring a response correct (for the simple reason that requiring subjects to make this further expansion would have involved more complicated instructions; the tendency to expand was not related to the tendency to do well, or to educational background). However, if the subject made an error in expanding, his response was called incorrect. Then only if a subject responds to the above stimulus:

A bird who lives in a house and he, the bird, is black
or A black bird who lives in a house
or a house for a blackbird

is he scored as having erred.

We distinguish four kinds of error:

a. errors of order (E_o): If the response the subject gave would have been correct had the order of the 13 to the 2 stress word in the stimulus been reversed, we say that the subject made an (O_1) error of order, e.g.:

Stimulus: bird^1 house^3 boot^2
Correct response: a boot you wear in a bird^1 house^3
Error of order: a bird^1 house^3 that has a boot in it
Stimulus for which this response would have been correct: boot^2 bird^1 house^3

Similiarly, when a single change of order within the 13 compound would make the response correct, this inversion was considered an (O_2) error of order, e.g.:

Stimulus: bird^1 house^3 boot^2
Error of order: a boot for house^1 birds^3, pet birds
Stimulus for which this response would have been correct: house^1 bird^3 boot^2

However, if the change of order involves inversion of one of the words joined under 13 stress, we consider that a different kind of error was made, e.g.:

Stimulus:	bird^{1} house^{3} boot^{2}
Response:	a house for birds'^{2} boots^{2}
Stimulus for which this would have been correct:	bird^{1} boot^{3} house^{2} or bird^{2} boot^{1} house^{3}

b. errors of stress (E_s): If a single change in stress would make the wrong answer the right answer, we say there was an error of stress, e.g.:

Stimulus:	dry^{1} house^{3} bird^{2}
Correct response:	a bird who lives in a dry^{1} house^{3} (or a dry^{2} house^{1})
Error of stress:	a house^{1} bird^{3} who is dry
Stimulus for which this would have been correct:	dry^{2} house^{1} bird^{3}

c. chaotic errors (E_{ch}): Where more than one change of stress or order or both would have to be made for the wrong answer to be a right answer, the error is called chaotic, e.g.:

Stimulus:	house^{2} bird^{1} black^{3}
Correct response:	paint for blackening house^{1} birds^{3}; as black as a house^{1} bird^{3}
Error of chaos:	a bird house that is black
Stimulus for which this would have been correct:	black^{2} bird^{1} house^{3}

d. grammatical error (E_g): If the subject fails to provide a nominal response at all, we call the error a grammatical error, e.g.:

Stimulus:	bird^2 wash^1 house^3
Correct response:	a house where birds are washed a laundry for birds
Error of grammar:	somebody is telling the bird to wash the house wash the house^1 bird^3! an Indian is saying: 'bird: wash house!'

In other words, it may be difficult to say just what a compound noun is, but we can assume it is not an Imperative.

3) discarding subjects: All subjects provided nominal and adjectival-phrase paraphrases most of the time (errors of grammar (E_g) were relatively infrequent). Ignoring E_g, then, we can see that there is a certain probability of providing a correct paraphrase by chance; if, for example, the 'test-word' is a noun, and the stress pattern provided is 132, the response pattern:

a 2 for 13
that is
etc.

is going to be called a correct response, while the other five possible arrangements of the nouns (represented above by their assigned stress names) are going to be called errors. However, since more than one kind of rule can generally be invoked in defense of other paraphrases, since the probabilities of chance correct responses are somewhat different for verb and adjective test words, since any word in any compound can be reassigned to another lexical class, and since, in fact, E_g does sometimes occur, it is impossible to provide any reasonable estimate of the probability of a chance correct response for any stimulus.

We felt safe in assuming, after considering the various factors above, that a subject who was correct over 25% of the time was not responding at random. Further, any subject who was reasonably rational in his responses to certain selected 'easy' compounds could be assumed to be attempting to follow the instructions.[1] Any subject who failed to meet one of these two criteria was discarded. On this basis, results for four of the eleven C-group subjects were excluded. It should be noted, however, that the discarded subjects did not in fact seem to be qualitatively different in their response characteristics from C-group subjects just ahead of them. All were sufficiently uniform in responding appropriately to the easy stimuli to conclude that they represented, so to speak, simply the bottom of the same curve. Nonetheless, we 'chopped off' this bottom of the curve to ensure that population differences could not be interpreted as failure of some few subjects to understand the instructions effectively.

---

[1] The chosen 'easy' compounds: black^2 bird^1 house^3; bird^2 wash^1 house^3, black^1 bird^3 house^2; glass^1 house^3 bird^2; thin^2 house^1 bird^3.

g. results:

1) gross effects:

(a) The gross effect of groups: The result that stands out immediately is an overwhelming difference among the three groups. The mean number of errors is 17.7 for Group A, 64.3 for Group B, and 89.1 for Group C. Figure 1 presents the individual error scores for each subject and shows that there is no overlap of groups A and C.

(b) Statistical method: It appears that, for at least some aspects of linguistic performance, there are individual and population differences. Given the dispersion of performance scores, we must adopt some techniques of data analysis that can handle some of the more troublesome consequences of this dispersion. We must, in other words, be able to cope with the problem of whether the effect of some variable is genuine or might be attributed to chance fluctuation. Some linguists may feel that this is regrettable; and perhaps it is. But it should be noted that we are not proposing a statistical model of language, or linguistic behavior, but only a way of extracting some linguistic events from the behavioral buzzing confusion in which these phenomena are surely embedded. It is not quite clear why linguistics should be offended by techniques that are

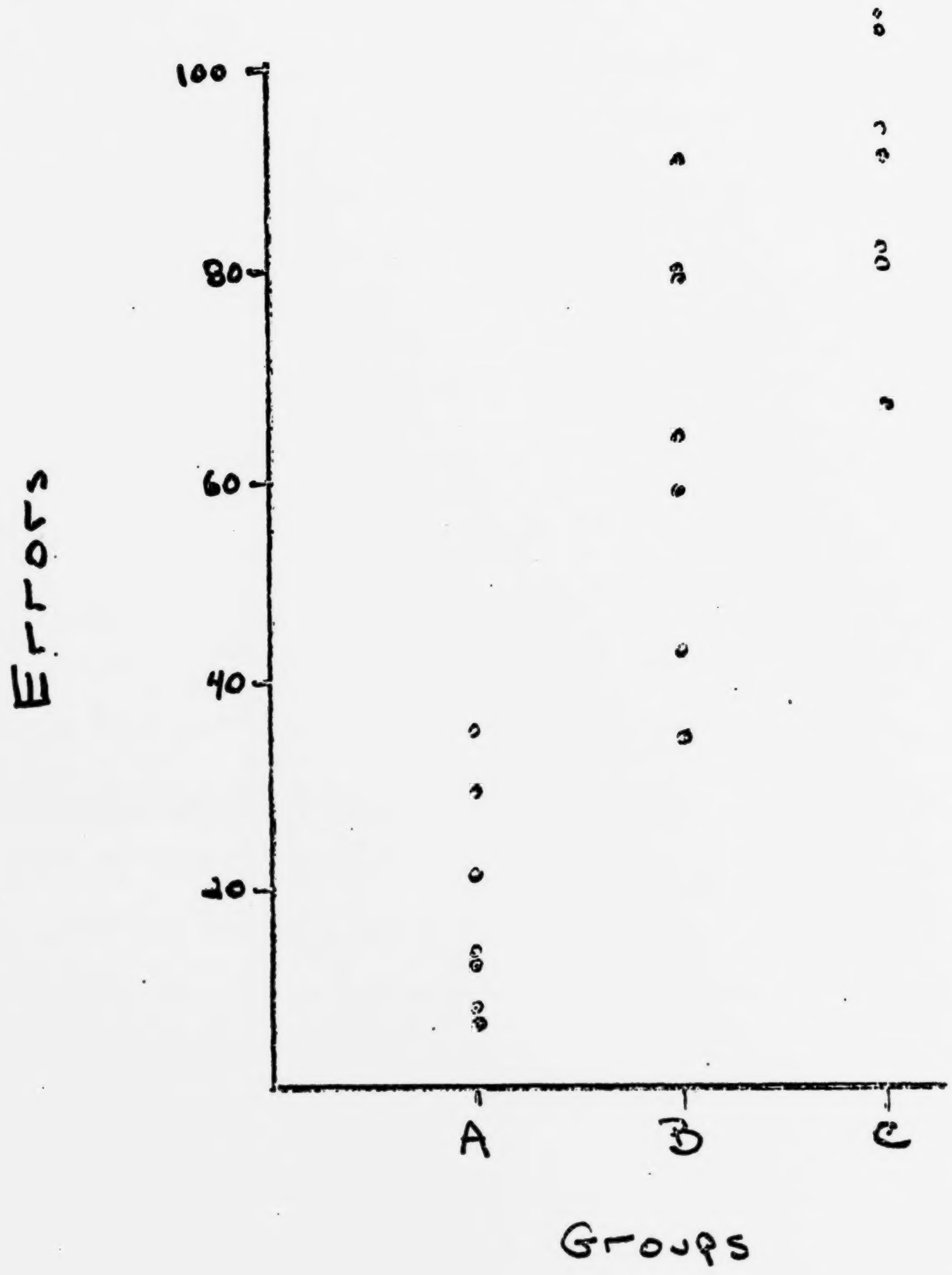

Figure 1. Distribution of error scores by group

routinely used in such varied disciplines as astronomy, genetics, agriculture, and psychology. At least some of these fields have not been the worse for their brush with the normal curve.

The primary statistical tool we will use is the analysis of variance. This technique, developed by Fisher (20), is particularly useful in multifactor experimental designs. Its main feature is a method of partitioning the variance and assigning it to various factors and the interactions among them. Thus for each factor examined, one can determine whether its effect reliably transcends chance fluctuation. This is accomplished by comparing the variance attributable to one or another factor with variance estimates of residual error. The analysis of variance is thus useful in separating out effects of the variables and in determining the contribution of each to the experimental results. Beyond assessing these "main effects", this statistical technique also allows us to determine whether the several factors produce their effects independently of each other, or whether the effect of one factor is different depending upon the value taken by the other. In the second case, we speak of an "interaction" between the two factors. Statistically speaking, this means that the difference due to a variation in factor A will differ depending upon factor B.

Technically, a number of preconditions have to be met to satisfy the assumptions of an analysis of variance. The variances of the subcells are assumed to be the same, the distributions normal, and so on. It is often possible to pre-test the data and, in case of nonhomogeneity of variance, to transform the data until the assumptions of the analysis are met (e.g. by changing *x* to log *x*, proportions to their arc sine equivalents). However, it then becomes necessary to justify the generalization from the result for transformed scores to the raw data. A series of studies (Norton, as reported by Lindquist, 49) has demonstrated that in practice the analysis of variance is remarkably insensitive to such violations of its assumptions, even where these are fairly extreme; we have therefore not attempted to correct for violations.

(c) Statistical treatment of gross effects: It is hardly surprising that the group difference in overall error scores is highly significant statistically. A simple analysis of variance shows an immense effect of groups [$F = 37.1$, *df* = 2 and 18, $p < .001$]. Further, Duncan's Multiple Range Test showed that all groups differed significantly from each other [$p < .01$].

It will be recalled that in addition to the population groups, subjects (within groups) were further subdivided into 'order-groups' that received the stimuli in three different serial orders. Some subjects began with a syntactically plausible and semantically familiar stimulus, others with a syntactically plausible but semantically odd stimulus, and still others with a syntactically opaque and semantically odd stimulus. This difference did not affect the final error-scores. An analysis of variance shows no trace of an order effect [$F < 1$].

It is nevertheless possible that there might still be an effect due to serial position. The question is: is an item simpler when it comes late in the series; in short, does the subject learn? To answer this question, we considered mean error scores for test-items 1-36; seven subjects received these as the first 36 items, seven others received them in position 96-132, seven others in position 109-144. The mean error scores do decline very slightly for groups that received the items later in the test, but the effect does not begin to reach significance. An analysis of variance for the 36 stimuli shows the usual huge effect of groups [$F = 42.2$, df = 2 and 12, $p < .001$], but neither the main effect of serial position nor its interaction with the group effect met anything like the usual criteria of statistical acceptability (both p values were larger than .25).

2. The relation of syntactic form to error:

In the previous analysis, we examined overall error scores for all 144 stimulus items. We did not ask what kinds of errors were made, or what kinds of stimuli led to errors. We turn now to these questions.

a. Categories of stimulus and response types:

1) Stimulus categories: Each stimulus item consisted of three words, two of which were always the same (bird and house), and a third that could be varied. We refer to the two constant terms as C, to the variable one as X. The first basis of classification of the stimuli is by the lexical classification of the X-term (see Table 2). For the moment, we restrict the analysis to X-terms that have unambiguous 'basic' lexical classification (Noun, Adjective, or Verb) in Table 2; excluded are stimuli containing stone, glass, dry, or shut as the X-term. The results for these ambiguously classified stimuli will be considered later. Their exclusion leaves 96 stimulus items.

Each X-term can occur in the 1st, 2nd, or 3rd serial position in the compound. We disregard, for the moment, the relative order of the two constant terms, thus for present purposes grouping house-black-bird with bird-black-house. The number of arrangements is doubled when we consider that for each of the three word-orders there are two possible stress patterns (*132* and *213*).

There are now three bases for stimulus categorization: (1) lexical class of the X-term; (2) position of the X-term; and (3) stress pattern of the compound. Thus there are 18 'phrase-categories'.[1]

2) Response categories: Just as we subcategorized the stimulus-types, we now distinguish among the error-types yielded b the scoring method (see Section). There are four kinds of error: E_o (errors of order)[2], E_s (errors of stress), E_{ch} (errors of chaos), and E_g (errors of grammar).

b. Conversion to ratio scores:

The preceeding analysis of group effects was based on the gross error scores for all subjects for all 144 stimuli. We are now asking about the distribution of the errors. Is it still reasonable to work with gross error scores? Suppose we want to know whether various phrase-categories are associated with certain kinds of errors. We would get our answer by summing up all of our subjects' error scores within each of the 72 subcells (18 phrase-categories and 4 error-types). But once we have done this, each error counts equally, whether it is made by a subject who errs very often or a subject who errs hardly at all. It is obvious that in the final reckoning, the result must be disproportionately affected by the subjects with higher error scores. Since we have already granted individual differences, we are no longer asking how many errors a subject made, but rather how these errors are distributed by phrase-category and error-type. We will therefore consider all subjects and consider all equally: subjects, rather than errors, will carry equal weight. This effect is accomplished statistically by a simple ratio transformation of the error scores: Each subject's error scores (within phrase-categories and error-types) were divided by the total number of errors made by this subject on the 96 stimulus items. It is obvious that the sum of these adjusted scores will now be equal (in fact, 1.00) for every subject; thus each subject contributed equally to the overall pattern of results.[1]

---

[1] Each stimulus-category is given a three-place name by (1) lexical class (N, A, or V); (2) position of X (1, 2, or 3), and (3) stress pattern. When the X-term has intermediate stress, the stress pattern is called positive (+); otherwise, negative (-). Intermediate stress occurs only in serial positions 1 and 3. In second position, X-terms with primary stress are called *E*, with tertiary stress, *D*. Thus N1+ includes *foot^2 bird^1 house^3*. N1- includes *foot^1 house^3 bird^2*; N2E includes *bird^2 foot^1 house^3*; N2D includes *bird^1 foot^3 house^2*.

[2] In this analysis, O_1 and O_2 errors are not distinguished. In experiments 2 and 3, this further distinction will be treated.

---

[1] By the ratio transformation, we have accorded equal status to all individuals, regardless of their differing propensities to err. This makes excellent sense statistically. It also is a more reasonable model of the linguistic behavior of the individuals and groups we are studying. Given the enormous differences in individual error scores, it is plausible to suppose that there are different 'baselines' for error, or general performance level, among subjects. If this is so, an analysis of variance on the 'raw' error scores will be very misleading. By analogy, suppose we are studying the effect of very hot weather on the performance of tennis-players; suppose further that one subject is a professional, and another is a rank amateur. As the temperature increases from 70 to 90 degrees, let us say, the professional's missed shots increase from 2 to 4 in 20. The amateur's increase from 8 to 16. If we use mean missed shots as our measure, and treat these data in their raw form, we will find an 'interaction' between temperature and professional status, i.e. the effect of temperature is not additively equal for the two subjects, since the professional's errors increased by only 2 while the amateur's increased by 8. It is more probable, however, that the effect of temperature is to double the number of missed shots regardless of the basic propensity to miss the ball. The interaction, then, is spurious, a result of the inappropriateness of the counting method. If, instad, we had originally transformed the scores to proportions (thus correcting for the differnt baselines), we would now find that the effect of temperature on mean adjusted error was additively equal, i.e. the differences between the differences were identical. Granting an overall difference of aptitude among our own subjects, it seems more likely that some factor of difficulty added to certain stimulus categories would increase the proportion of error more or less equally, than that it would increase the number of errors per subject by some constant amount. Without a model that actually predicts the relation of syntactic (and semantic) form to error production, it is difficult to speculate much further than the ratio model we have tentatively accepted here.

But it should be pointed out that there are drawbacks to a ratio analysis that are independent of the plausibility of the model for these data. The A-group subjects make far fewer

c. Analysis by group, phrase, and error type: Here we consider the relations among three main factors: Groups (A, B, C), phrase-category (the 18 types described above), and error-type (E_{∂}, E_s, E_{ch}, E_g). An analysis of variance was performed based on the mean adjusted error scores of each subject for each of the 18 phrase-categories. Absolute differences between groups have of course been eliminated by the score adjustment.[1]

Table 3 presents mean errors by group, phrase-category, and error type. The most casual inspection indicates that there are wide variations resulting from phrase category, and that these are not distributed evenly over types of error.

Footnote continued:

errors, thus yielding data that are less reliable, especially when the results are broken up into numerous sub-categories. For these subjects, one random error (and even A-group subjects are prone to random lapses of attention) may have a disproportionate statistical effect. There is no perfect way of balancing the advantages of the absolute and the ratio analyses. It is worth noting that, for the most part, the analyses led to rather similar results whether carried out on the raw scores or on the transformed ratio scores.

[1] A simple analysis of variance using raw scores was performed for the 96 stimuli considered here (the ambiguously classified stimuli have been removed) to determine whether the group-effect for all 144 stimuli was still present. The same massive effect was found [$F = 28.37$, df = 2 and 18, $P < .001$].

Table 3. Mean adjusted errors by group, phrase-category, and error-type (adjusted errors x 1000).

Phrase-category	Errors of Order			Errors of Stress			Errors of Chaos			Errors of Grammar		
	A	B	C	A	B	C	A	B	C	A	B	C
N1+	9	0	14	0	23	14	0	7	10	0	0	12
N1-	10	13	11	0	36	14	0	3	3	0	0	3
N2E	0	22	9	0	19	13	0	6	6	0	0	0
N2D	0	8	7	15	10	34	0	7	6	0	0	0
N3+	0	12	3	10	0	6	0	3	3	0	7	0
N3-	38	36	15	0	42	48	0	16	7	0	0	0
A1+	0	0	4	0	0	12	0	0	2	0	3	0
A1-	0	10	5	96	52	70	0	0	0	0	0	0
A2E	26	23	12	10	17	8	6	10	10	24	0	0
A2D	36	19	9	0	33	55	0	10	31	0	0	2
A3+	18	68	79	10	0	0	0	10	5	0	0	0
A3-	164	42	32	16	0	3	44	64	58	0	0	2
V1+	140	8	4	0	9	15	0	7	2	89	25	53
V1-	25	31	18	46	35	8	0	10	18	10	17	35
V2E	19	16	6	10	3	8	6	24	18	0	7	16
V2D	46	5	0	0	13	10	6	21	19	0	4	15
V3+	0	22	17	30	9	2	0	22	10	0	2	16
V3-	34	47	20	6	7	22	0	18	23	0	6	10

To disentangle the various effects, let us first simplify by lumping results for all population groups. This is done in Table 4, which shows mean number of errors as a function of category. A consideration of total number of errors (Col. 5, Table 4) shows that some categories are fairly easy to deal with (e.g. A1+, N3+, N2E), while others are very hard (e.g. A3-, V1-, A3+). It thus appears that there is a difference in the difficulty of the various categories, quite apart from the kind of errors they produce. This main effect of phrase-category is highly significant [F = 7.87, _df_ = 17 and 1,490, p < .001].

There is similarly a significant main effect of error-type [F = 23.4, _df_ = 3 and 1,388, p < .001]. Errors of order and stress occur more often than errors of chaos and grammar.

That this last main effect has no independent interest is shown by the fact that phrase-category and error-type interact; thus by manipulating the phrase-categories, we can affect the relative frequency of various kinds of error. Errors of grammar occur most frequently when the X-term is a verb (particularly in first position), errors of order with postnominal adjectives (A2D, A3), and errors of stress with noun X-terms. This interaction of phrase-category with error-type is significant [F = 2.79, _df_ = 51 and 1,490, p < .01], and has certain obvious syntactic correlates that we will discuss later.

Table 4. Mean adjusted errors by phrase-category and error-type (adjusted errors x 1000).

	Errors of Order	Errors of Stress	Errors of Chaos	Errors of Grammar	All Errors
N1+	8	12	5	4	29
N1-	11	17	2	1	31
N2E	10	11	4	0	25
N2D	5	19	5	0	29
N3+	5	5	2	2	14
N3-	30	30	8	0	68
A1+	2	4	1	1	8
A1-	5	73	0	0	78
A2E	20	12	9	8	49
A2D	21	29	13	1	64
A3+	55	3	5	0	63
A3-	79	6	55	1	141
V1+	51	8	3	56	118
V1-	25	30	9	21	85
V2E	14	7	16	7	44
V2D	17	8	15	6	46
V3+	13	14	10	6	43
V3-	34	12	14	5	65
All phrase-categories	405	300	176	119	1000

We now ask how group-membership interacts with the effects of phrase-category and error-type. Table 5 presents mean adjusted error scores for groups A, B, and C for each of the four error-types. Of considerable interest is a significant interaction between group-membership and error-type [$F = 2.60$, df = 6 and 1,490, $p < .02$]. The A-group makes proportionally more errors of order, and fewer errors of chaos and grammar; the reverse is true of the C-group, while the B-group falls in between.[1] Referring back to Table 3, we can attribute this effect primarily to two very difficult phrase-categories (A3- and V1+). In both these cases, the A-group responds with E_o. For A3-, Groups B and C give E_{ch}. For V1+ group B and especially group C give mainly E_g.

While there is an interaction between group and error-type, there is little evidence of an interaction between group and phrase-category [$F = 1.16$, df = 34 and 1,490, $p > .05$]. It appears that phrase-categories that are difficult for the linguistically inept are difficult also for those who are more facile; the absolute level of difficulty (i.e. the baseline for error) varies, but the relative order is pretty much the same. In a way this is reassuring: granted that there are some major differences among individuals (the raw-score effect of groups), and differences in how they cope with difficulty (the interaction of groups with error-type), at least they all speak the same language (there is no interaction of groups with phrase-category).

---

[1] This effect can be enhanced by use of a more refined scoring technique which will be explained later (App. B).

d. The relations between phrase-category and number and kind of error: The preceeding analysis indicated that the phrase-category affects the number of errors committed, and that different phrase-categories are associated with different kinds of errors. Several further analyses were undertaken to examine these effects more closely.

1) The effect of lexical class: Inspection of Table 4 suggests that one determinant of difficulty and of kind of error is the lexical class of the X-word. To test this suggestion more directly, we grouped the 18 phrase-categories by the three kinds of X-term (Noun, Adjective, Verb).

The effects of lexical class on performance are shown in the bargraph of Figure 2. Recall that E_o is much the most frequent error-type committed. Figure 2 shows that for both Nouns and Adjectives, E_o occurs more often than any other

Table 5. Error-type as a function of group. (Mean adjusted errors x 1000)

	Group:		
Error of:	A	B	C
Order	566	381	266
Stress	248	310	340
Chaos	62	238	231
Grammar	124	71	163

error. But E_s is very much greater for Nouns than for Adjectives, and E_o is greater for Adjectives than for Nouns. E_g is almost absent except with Verb X-terms. We reserve for later discussion an examination of what these associations (and also the scatter of E_{ch} over lexical classes) imply. An analysis of variance showed a significant main effect of lexical class [$F = 10.85$, df = 2 and 1,457, $p < .001$]: Nouns lead to fewer errors than either verbs or adjectives. More important, there was a significant interaction between lexical class and kind of error [$F = 4.44$, df = 6 and 1,457, $p < .01$]. Nouns are associated with errors of stress, adjectives with errors of order, and verbs with errors of grammar.

2) The effect of position: The very powerful effect of phrase-category (on both number and kind of error) can be accounted for in part by the lexical class of the X-term, as the previous analysis has shown. But it is quite obvious that a good deal of variation is still left unaccounted for. Some of the variance that remains is clearly attributable to the position of the X-term. For example, phrase-categories with Adjectives in first position lead to little error, while phrase-categories with Adjectives in last position are very difficult. Such effects are immediately apparent upon inspection of Table 3; they should be

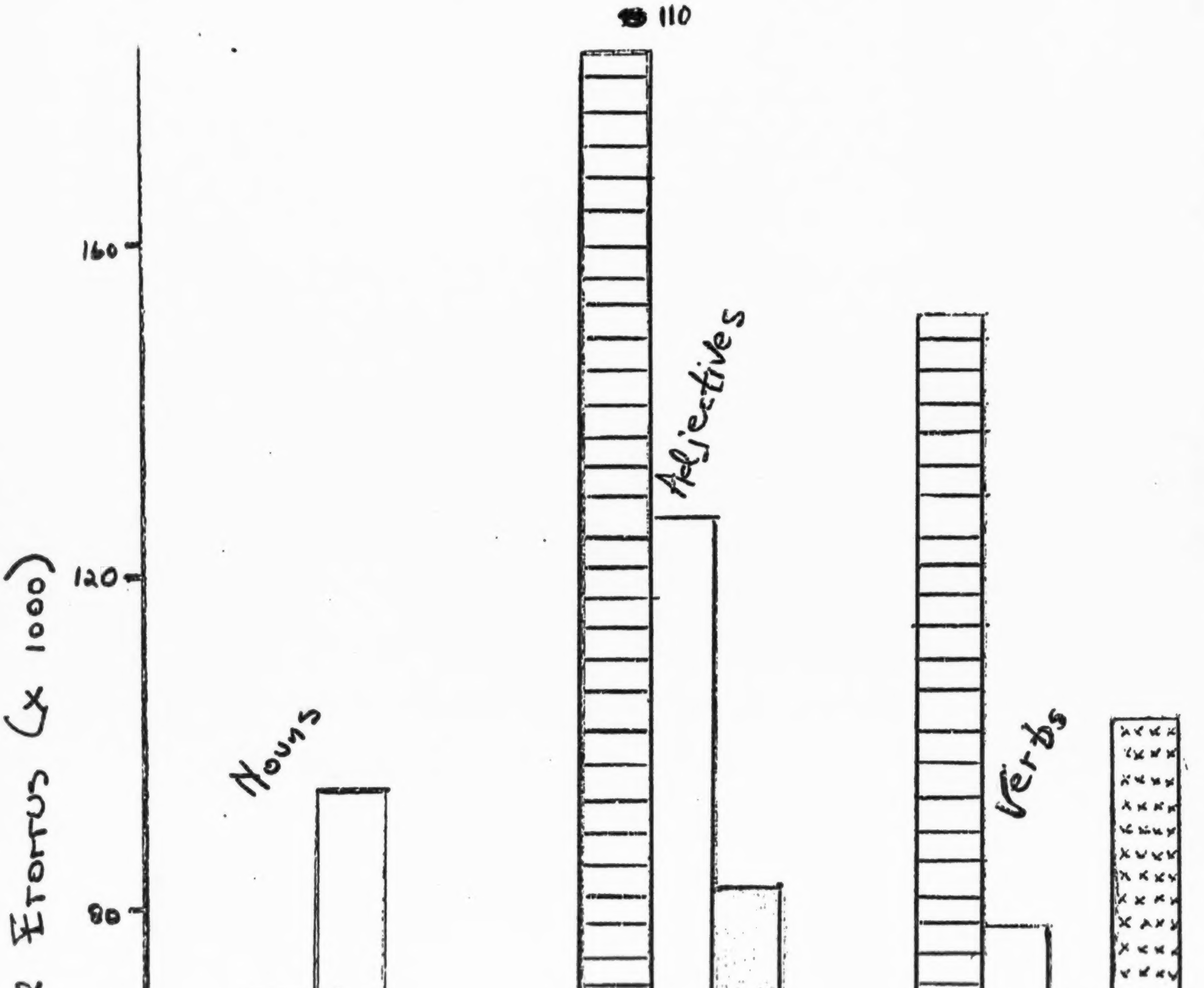

Figure 2. Error type and lexical class.

even more apparent upon inspection of the English language: Adjectives often begin phrases, but rarely end them. Given the enormous importance of sequential patterning in English, the fact that lexical class must interact with position need hardly be demonstrated. (As an exercise in pedantry, we did perform the statistical demonstration, performing an analysis of variance on overall adjusted error scores as a function of lexical class and position. We are happy to announce a highly significant interaction between lexical class and position of the X-term [F = 6.82, df = 4, and 1,482, $p < .001$]).

3) The effect of 'unity': The differences among the phrase-categories are not, however, abolished even if we (statistically) equate for lexical class, the position of the X-term, and the interaction between them. A major source of variation still remains. Inspection of Table 3 suggests that this factor is connected with the stress pattern, especially in case the X-term is in first or third position. Several further analyses were undertaken to isolate this factor and see how it operates.

Consider stimulus items with serial order XCC and CCX. Two stress patterns (132, 213) occur with both these serial orders in our stimuli. The stress patterns have the effect either of 'unifying' the two constant (C) terms, (i.e., $C^1C^3X^2$, $X^2C^1C^3$), or 'disunifying' them (i.e., $C^2C^1X^3$, $X^1C^3C^2$). In other words, the 13 stress subpattern (whether it preceeds or follows the secondary (2) stress) conjoins two words into a subcompound (e.g. bird^1 house^3).

Suppose there is a tendency for subjects to 'expect' that the two C-terms will be a subunit (i.e. will have the 13 subpattern). Given the C-terms we used, this would not be surprising. Bird^1 house^3 is a familiar unit, known to speakers as a lexical item; house^1 bird^3 is not familiar, but it requires no great flight of the imagination to connect it, by analogy, with known items (e.g. jail-bird, house-cat, field-mouse, barn-owl). At any rate, neither of these pairs, when 'unified', violates cooccurrence restraints in the language or semantic plausibility. When, on the other hand, a C-term is unified with an X-term, a cooccurrence restriction may or may not be violated (e.g. black^1 bird^3 and wash^{1}house^3 are familiar and plausible; bird^1 boot^3 and dry^1 house^3 are unfamiliar but plausible; house^1 foot^3 and bird^1 thin^3 are unfamiliar, implausible, and just possibly violations of real cooccurrence restraints in English). We would expect fewer errors where cooccurrence restrictions are respected. Since for 'unified' CC, there is never a violation, and for 'nonunified' CC, there is often such a violation, unified CC ought, on average, to be easier. Notice that if CC is non-unified, and in fact CX or XC does violate a cooccurrence restriction, this phenomenon might be further enhanced: not only must the subject deal with an eccentric semantic event, he must take care not to be beguiled by a normal semantic event that he can bring into existence by just a wee flim-flam with the stress pattern. Presented with foot^1 bird^3 house^2 (clearly a home for livery birds) one might be tempted to respond 'a basement apartment for birds', although that is a foot^2 bird^1 house^3, as we know, and, in all fairness, not an altogether unextravagant alternative.

To determine whether such a phenomenon actually exists, we compared mean adjusted error scores and error-types obtained in which the C-terms were unified with those obtained for items where the C-terms were not unified. For these purposes we of course do not consider items in which the X-term is in second position (where the C-terms cannot be unified). Again, we exclude the lexically ambiguous stimuli. Then there are 64 stimulus-items for the following analyses.

Table 6 presents mean adjusted scores for unified and nonunified stimulus items by error-type. As the table indicates, the stress difference indeed produces enormous differences in the tendency to err. The number of errors for items with nonunifying stress is almost double that for items with unifying stress. As expected, the main effect of the unity factor is highly significant [$F = 11.25$, df = 1 and 964, $p < .001$].

Table 6. Errors as function of stress. (Mean adjusted errors x 1000)[1]

	Error-type				All errors
	Order	Stress	Chaos	Grammar	
Unified stress	133	47	27	69	276
Non-unified stress	184	167	88	28	467

[1] Figures do not total to 1000 since they only include data from X-term positions 1 and 3.

If the unity factor can in fact be interpreted as we say, then the errors produced by this stimulus factor should be primarily errors of stress and chaos.[1] As Table 6 suggests, this turns out to be so: The unity factor interacts significantly with error-type [F = 5.88, df = 3 and 964, $p < .01$], and the interaction is attributable to a sharp increase in E_s and E_{ch}.

Figure 3 graphically presents the unity factor as a function of groups. There is no trace of an interaction between unity and group membership, ($F < 1$). Apparently, A-group subjects are as prone to such interference (in proportion to their baseline for error) as are C-group subjects.

4) The effect of familiarity: It is possible that overall difference in the familiarity of the subcompound (the item under 13 stress) may affect the tendency to err. We have mentioned above a very obvious difference between the well known compound bird-house, and the (probably) unknown, though plausible, compound house-bird. House-bird can be understood either by reference to the compounding rules of the language or by some other more direct

---

[1] It will be recalled (see Section IIIB2f) that an error of chaos is said to occur when a response reflects both an error of stress and an error of order. Since we therefore expect some covariance of E_s- and E_{ch}, we here expect a rise in both.

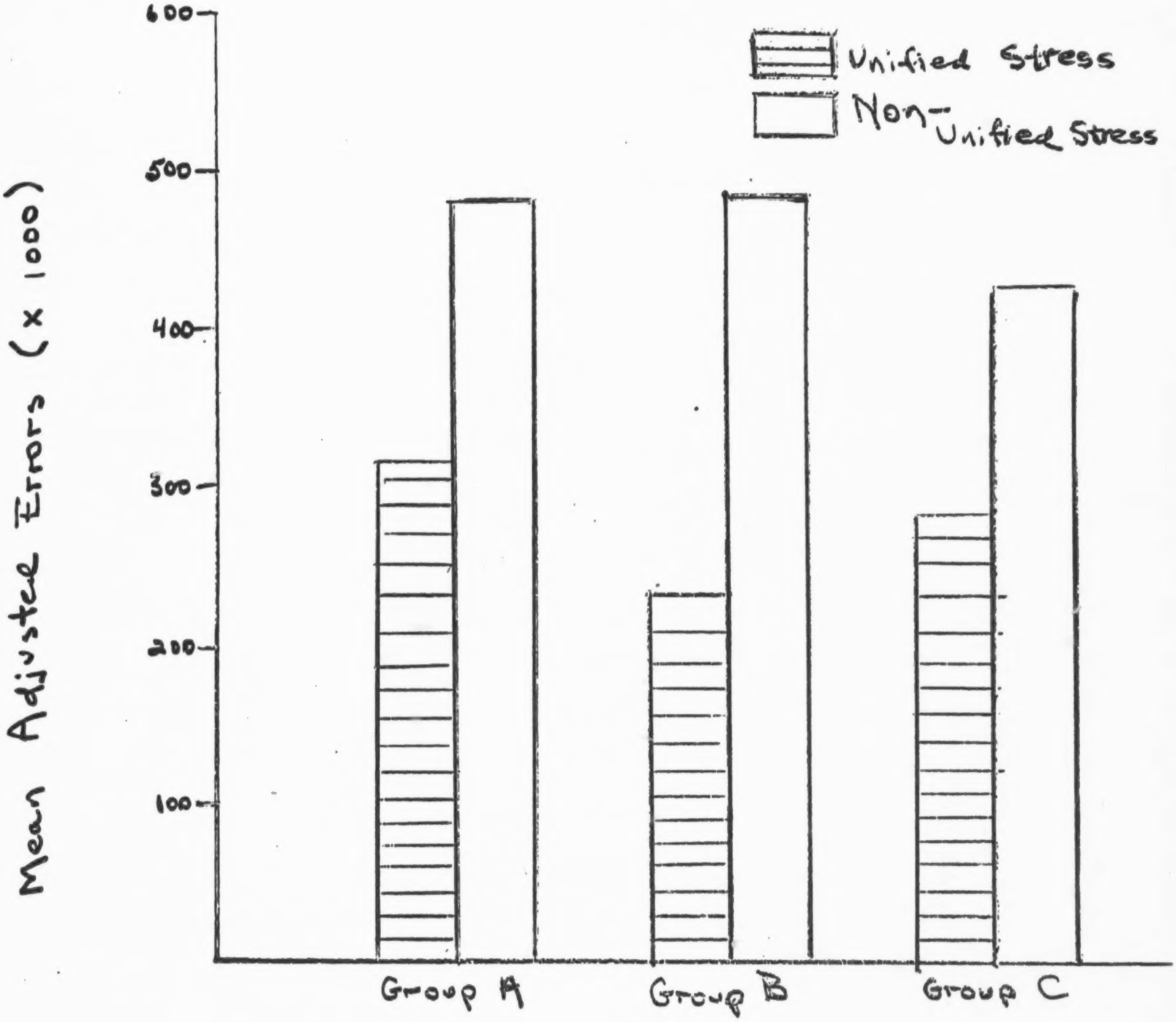

Figure 3. The unity factor by group.

mode of analogy (e.g. to *house-cat*). But a bird-house is just a nest. It is known as a lexical item; if you also know it is called a *bird-house* because it is a house for birds, so much the better, but this step is not necessary.

The 'computation' or comprehension of three word compounds containing *bird-house* might then be expected to be somewhat easier than the comprehension of those containing *house-bird*. Should this mean that, in our test, the sequence *bird*, *house* will lead to fewer errors than the sequence *house*, *bird*? Not necessarily. We already know of the power of the unity factor. What we expect is that familiarity should facilitate a correct response if the stress pattern demands that the two C-terms be responded to as a unit ('+' stimuli); but if the stress pattern is such ('-' stimuli) as to disunify the C-terms (i.e. demand that they be taken as non-units), then the effect of familiarity of the compound should pull against the demands of the test situation. If *bird-house* is a better unit than *house-bird*, by the same token it ought to be a worse nonunit (e.g., *bright*^{1} *bird*^{3} *house*^{2} ought to be harder than *bright*^{1} *house*^{3} *bird*^{2}). In short, we expect that the two factors of familiarity and unity will interact. Given unifying stress, familiarity should help; given disunifying stress, it should hinder. If both effects are about equally strong, there should be no main effect of familiarity, for it will be cancelled out.

To test these speculations, the same 64 items were considered again, but now grouped by both familiarity of compound and by stress pattern. The results are presented in Figure 4. As the figure indicates, the two factors did interact and in the expected direction. An analysis of variance shows the interaction to be highly significant [F = 11.69, df = 1 and 53, p < .001]. The same enormous effect of unity already discussed was found again [F = 20.3, df = 1 and 53, p < .001], but there was, as expected, no trace of a main effect of familiarity [F < 1]. It is worth noting that the interaction between familiarity and unity occurred rather uniformly for all three groups. There was no triple interaction (G x F x U), nor any simple group by familiarity interaction.

e. Mixed lexical items: We have thus far excluded from analysis stimulus-items containing the words glass, stone, dry, and shut, which we said (Table 2) had more than one 'basic' lexical classification. We turn now to an examination of these mixed items. In all, we are dealing with 48 stimulus items.

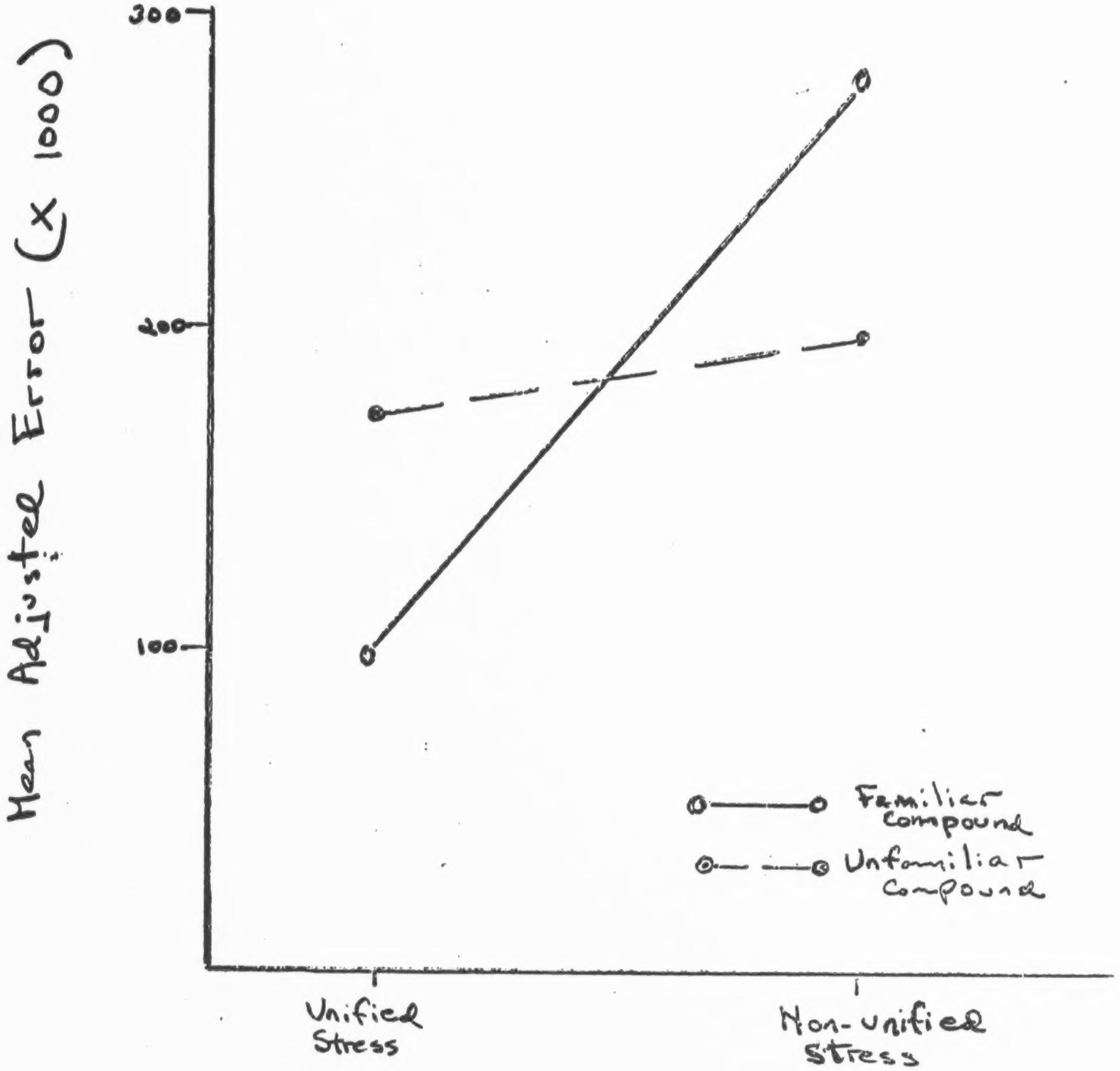

Figure 4. The effect of familiarity of compound under different conditions of stress.

1) Stimulus subcategorization: Without reawakening the argument of whether *stone* and *glass* are indeed adjectives (see footnote 1, p. 81), we can agree that words like these are comfortable in prenominal position with phrasal (*21*) intonation, while other nouns are not (e.g., *stone^2 house^1* vs **dog^2 house^1*). In this sense only we say that *stone* and *glass* are 'mixed', having positional privileges in common with both adjectives (as in *glass^2 house^1*) and nouns (as in *house^1 glass^3*); we group these two words into a category "AN". *Dry* and *shut* have the positional privileges of both verbs and adjectives; we group them as "VA".

2) Comparison of mixed and simple lexical items:

(1) Gross error: Our initial supposition was that stimuli with mixed X-terms ought to lead to less error than stimuli with simple X-terms. After all, it is difficult to deal with an adjective in third serial position, and easier to deal with a noun in third position. This being so, an adjective that may alternatively be viewed as a noun ought to present less of a problem than a simple adjective.

To test this notion, we compared the mean number of errors when the X-terms were simple (that is N, A, or V) and when they were mixed (AN or VA). These means are presented in Table 7. As the table shows, our initial notion may have been plausible, but it turns out to be false: mixed X-terms

Table 7. Mean gross errors for simple and mixed X-terms.[1]

Error-type	Lexical class of X term				
	N	A	V	AN	VA
Order	1.0	2.2	1.4	1.5	2.2
Stress	1.6	1.7	.9	1.8	1.6
Chaos	.5	1.3	1.2	.7	1.2
Grammar	.1	.[illegible]	1.6	.0	.7
All errors	3.2	5.2	5.1	4.0	5.7

---

[1] Mean gross errors per X-term.

do not lead to fewer errors than simple ones. The AV X-terms lead to more errors than either the A or the V X-terms. (However, neither the difference between A and AV nor that between V and AV achieves statistical significance; for the first, $t = .94$; for the second, $t = 1.41$. With $df = 20$ in both comparisons, $p > .10$). The AN X-terms lead to a level of difficulty midway between that of N X-terms ($t = 2.52$, $df = 20$, $p < .01$) and that of A X-terms ($t = 2.52$, $df = 20$, $p < .01$).

These effects held for all groups. Consider the AN items, which had a mean error score midway between that obtained for A and N items. In all, 14 subjects found these items harder than those having N X-terms; of these 14 subjects, five were in group-A, four in group-B, and five in group-C. On the other hand, only three subjects found these AN items harder than items with A X-terms; of these three subjects, one came from each of the groups. The pattern is similar for the VA items. Only seven subjects (three from group-A, one from group-B, three from group-C) made more errors on A than on AV items; only eight subjects (three from group-A, one from group-B, four from group-C) made more errors on V than on AV items.

(2) Error-type: We already know that cer- error-types are associated with the different lexical classes

of the X-terms. Errors of order are associated with adjectives, errors of stress with nouns, and errors of grammar with verbs (see Figure 2). Perhaps by looking at the kind of error made for the stimuli with mixed X-terms, we can discover how they are interpreted by the subjects.

Inspection of Table 7 suggests that AN words were, overall, interpreted as nouns. However, in third position (where -- of all places -- it is dangerous to interpret them as adjectives) they were surprisingly often taken as adjectives, even though this rendered them artificially difficult. Similarly for VA, there are many instances in which the mixed X-term is evidently viewed as an adjective when that would be most difficult to interpret (in third position), and as a verb when _that_ would be most difficult (in first position), thus increasing both errors of order and errors of grammar beyond the number for simple V or A. It is as though the subject, rather than looking for ways of being correct, was seeking out ways of getting to a wrong answer.

3. Experiment 2

We have seen that there are gross differences between the behavior of more educated and less educated populations in responses to this kind of test. It is highly improbable that there are social-class or educational-class dialect differences that bear on the structure of compounds of the kinds we studied.[1] Then it is likely that our results reflect global differences in the handling of linguistic materials by recognizable populations within the same dialect group.

In reporting the results of Experiment 1, we assumed that the less educated groups in fact understood the instructions. The assumption seems warranted because (1) we discarded subjects whose scores were not overwhelmingly beyond chance expectations (see Section III B3); and (2) responses of all subjects -- regardless of baseline for error -- were found to be systematically related to phrase-category in the same way, i.e. all subjects seemed to be responding similarly to the same features of the stimulus situation.

Yet even if all subjects did understand the instructions, it is possible that they differed in 'imaginativeness' or 'ingenuity' in dredging up semantically plausible interpretations for some of the compounds. There is no doubt that since a large number of the compounds were semantically bizarre, a large number of correct paraphrases had to be bizarre also. Ingenuity aside, perhaps the less-educated subjects were uncomfortable (at being wrong, at seeming foolish) so they did not verbalize semantically odd responses.

---

[1] The social or dialect-group interpretation of the results will be shown to be unacceptable in Experiment 3 in which we finally do find a few secretaries whose performance is fairly close to that of the graduate students.

Furthermore, the inability to generate paraphrases spontaneously need not necessarily bear on the question of whether the subjects understood the phrases. Although our intuition was that the measure of understanding is 'being able to say it in one's own words', that requirement might well be too strong.

For these reasons, we had to consider whether the task had been well designed to display the subjects' maximal comprehension. We therefore designed a forced-choice experimental situation in which the subject's task was to indicate his preference between two proposed paraphrases. Since the subject is now asked merely to choose between equally bizarre alternatives, the issues of verbal imaginativeness and temerity are removed, and the task becomes in some sense a more passive one.

a. Subjects: Nine months elapsed between the first experiment and this one. Six of the graduate students and three of the secretaries were no longer available, so some new subjects were added. Thus for some subjects, nine months to a year elapsed between tests, while for the new subjects only a few weeks elapsed between tests. This large difference in test-interval was found to be irrelevant to the subject's consistency in performing the two tasks. The college group ("B") was not retested.

Subjects were:

Five graduate students from the population previously tested.

Twelve secretaries from the population previously tested.

Results for two secretaries were dropped from the analysis because these subjects failed to provide relative-clause responses frequently enough for us to believe they understood the instructions (see p. 92). We no longer, however, required a minimum score of 25% correct, because internal features of poor subjects' responses seemed no different from those of better subjects. Two secretaries(S4 and S8 in this new group get only 20% of the stimuli right in the original task.

b. Materials: The same taped list of stimuli was used. Two decks of response-cards were developed. The first listed the modal correct response of the new A-group to each of the items in the spontaneous generation task. The second listed the modal error (in most cases this was the modal response of the C-group) of the new C-group for each of the original items.

The two decks were randomized, using a random number table, so that the 'correct' response now might appear either in the left-hand or the right-hand pack.

c. Instructions: The subject was reminded of the earlier task, and told that this time she was asked merely to choose the card that 'meant most nearly the same thing' as the compound she heard. Stress was indicated on the card by hyphenation, as in ordinary English orthography, and the subject was taught and pretested to assure that she understood how to pronounce the responses listed on the cards; this seemed to create no difficulty. The instructions from the spontaneous-generation task were now reread to the subject, with appropriate changes for the new task.

d. Method: The two randomized decks were laid face-up before the subject (i.e. the correct and the incorrect responses to stimulus 1 were first exposed). The stimulus was then played on the recorder, and the subject handed to the experimenter the card she preferred. As in the previous experiment, replay of the stimulus was allowed at the subject's request. The experimenter now removed the card that had not been chosen (thus exposing the choices for the next stimulus), recorded the response, and played the next stimulus.

b. results:

1) Gross effects: Our first question is perhaps the most important here: Did the subjects do better when their task was to recognize the correct alternative rather than to produce it? If they did markedly better, we would suspect that the results for Experiment 1 were artifacts of the procedure. As it happens, subjects' behavior was substantially unaltered.

Table 8 presents the percentage of correct responses for each of the subjects in Experiment 1 ("generation") and Experiment 2 ("choice").[1] At first glance, the table seems to suggest that there has been a marked improvement for all but one C-group subject.[2] But this improvement results from an artificial inflation of the recognition score by chance factors. After all, the subject now has only two alternatives: if she did not understand a word of English, she would come up with a score of 50% "correct", a fact well known to all who have taken true-false objective examinations. We must obviously correct for this chance factor.

---

[1] N is less than 144 for most subjects because we exclude trials on which the subject's spontaneous paraphrase was not intelligible on the tape.

[2] For A-group subjects there is also perhaps an improvement, but an obvious ceiling effect makes it difficult to evaluate this result.

Suppose we had some means of detecting how many of the correct choices made by the subject are "genuine", i.e. reflect "what she really knows". Let us call the number of genuine correct responses M. If there are N stimuli altogether, there are of course N - M cases for which the subject does not "really know" what she is doing. Given that she does not know, she can do no better than guess. Under these circumstances the chances of guessing correctly are .5; hence 1/2 (N - M) "correct" choices will be added to his score. The total number of correct choices (T), genuine and chance, is thus:

$$T = M + 1/2\ (N - M)$$

Can we find some way of estimating M? In fact we can, for we have the number of correct paraphrases given by the subject in Experiment 1. The results of that experiment suggest that the probability of generating a paraphrase correctly by chance is essentially zero. Substituting appropriately we can now derive what T should be on the assumption that the subjects perform identically on both tasks. These predicted T-values are presented in the last column of Table 8. It is clear that, once the choice scores are corrected for chance, Experiment 1 leads to more correct responses than does Experiment 2. This outcome is the more surprising because we would really have assumed subjects would be correct more than half the time even for

Table 8

Group	Subject	N	Generated correct	Obtained chosen correct	Expected chosen correct
C	S1	143	56	77	99.5
C	S2	141	78	94	109.5
C	S3	143	60	88	101.5
C	S4	142	26	63	84.0
C	S5	143	59	89	101.0
C	S6	117	41	68	79.0
C	S7	140	84	82	112.0
C	S8	125	28	58	76.5
C	S9	144	57	79	100.5
C	S10	135	50	81	92.5
A	S11	144	138	138	141.0
A	S12	144	110	136	127.0
A	S13	144	128	137	136.0
A	S14	144	126	133	135.0
A	S15	144	137	138	140.5

Comparision of number of correct responses for the "generation" and "choice" conditions stimuli which they had not generated correctly, because the correct paraphrase -- which the subject might not previoulsy have 'thought of' -- is presented for consideration.

Since Experiment 2 did not improve the performance of group-C subjects, we can reject the hypothesis that their ineptitude in generating paraphrases was an artifact of the procedure. But they did not merely fail to improve, they seem to have gotten worse. Why?

2) Choice performance as a function of the alternatives: The number of predicted correct choices (T) was estimated on the assumption that when the subjects did not really understand the compound, they would respond by chance. The fact is that they did not do so.

Table 9 displays the results in terms of the item-by-item consistency of each subject in the two experiments, i.e. whether the subject chose correctly just in case he had generated a correct paraphrase for a given item. It is obvious from this table that the subject is least likely to choose wrong if she had generated a correct paraphrase. But does the subject *prefer* to choose wrong if she had generated an incorrect paraphrase? This matter is unclear from Table 9: six out of ten C-group subjects seem to show this preference, while four go in the opposite direction.[1] For the A-group, the number of wrong responses is too small to support any conclusions, but they appear in general to choose correctly even when they had generated unacceptable paraphrases.

In order to clarify the question of whether subjects in fact prefer to choose the paraphrases they had previously generated, we can sensibly consider only those instances in which either the correct paraphrase or the error available as choices was equivalent to the paraphrase actually generated by the subject in Experiment 1, i.e. we exclude now all trials in which the subject did not have the alternative of effectively repeating her original response.[2]

---

[1] The ambiguity of these results compared with the model tested in Table 8 implies that that model was in error in assuming that the probability of a chance correct spontaneous paraphrase was close to zero. Lenient scoring and the benefit of a doubt in constructionally ambiguous responses no doubt account for this fact.

[2] Judging whether the subject's response was in fact available was a difficult matter, for we could not require word-by-word identity. For example, *a bird who lives in a black house* and *a bird who lives around black houses* were *considered* identical for these purposes. However, *a bird of the species "black-house"* was considered a different answer. *All questionable cases* were excluded. Judgments were made without knowledge of the subject's actual forced-choice response. The subject's responses were available as choices between 42% of the time (S9) and 81% of the time (S13).

Table 9

Group	Subject	N	Gen. correct Chose correct	Gen. correct Chose wrong	Gen. wrong Chose correct	Gen. wrong Chose wrong
C	S1	143	41	15	36	51
C	S2	141	56	22	38	25
C	S3	143	44	16	44	39
C	S4	142	14	12	49	67
C	S5	143	49	10	40	44
C	S6	117	26	15	42	34
C	S7	140	64	20	18	38
C	S8	125	13	15	45	52
C	S9	144	31	26	48	39
C	S10	135	43	7	38	47
C	Total	1373	381	158	398	436
A	S11	144	133	5	5	1
A	S12	144	109	1	27	7
A	S13	144	123	5	14	2
A	S14	144	119	7	14	4
A	S15	144	135	2	3	4
A	Total	720	619	20	63	18

Consistency of subjects between test and retest

Table 10 consists of contingency tables for C-group subjects, indicating that they tend to choose as they had generated. Each subject shows this same effect, seven of the ten significantly so. The sum of chi-squares for all C-group subjects is highly significant, yielding a chi-square of 113.9 which with df = 10 is significant well beyond the .001 level. Very clearly, subjects tend to perservere in their correct responses. For eight of the ten subjects, the tendency also to persevere in errors enhances this result, and only one subject (S2) shows a weak effect in the other direction. Over all C-group subjects, the probability of choosing a modal error in Experiment 2 if it was generated in Experiment 1 is considerably higher than .5. In table 11, the responses of A-group subjects are also indicated as contingency tables, although the paucity of errors does not allow us to make a realistic statistical test of significance. Notice, however, that the apparent tendency of A-group subjects to choose correctly even when they had generated incorrectly is (Table 9) minimized, just as it is for the C-group, when we consider only cases when the subject had his own generated response available as a choice: apparently, it is other people's errors that subjects correct, but they are satisfied to persevere in their own errors if those are available. It seems safe to conclude that it is not the case

Table 10

S1

	R	W
R	33	17
W	9	31

$\chi^2 = 16.90$

$p < .001$

S2

	R	W
R	35	18
W	11	16

$\chi^2 = 3.63$

$.10 < p < .05$

S3

	R	W
R	37	31
W	12	32

$\chi^2 = 7.94$

$p < .01$

S4

	R	W
R	11	16
W	7	33

$\chi^2 = 5.49$

$p < .02$

S5

	R	W
R	45	18
W	9	32

$\chi^2 = 24.36$

$p < .001$

S6

	R	W
R	19	11
W	9	18

$\chi^2 = 5.11$

$.02 < p < .05$

S7

	R	W
R	34	5
W	10	26

$\chi^2 = 27.23$

$p < .001$

S8

	R	W
R	11	22
W	11	33

$\chi^2 = .64$

S9

	R	W
R	23	10
W	18	10

$\chi^2 = .20$

S10

	R	W
R	35	15
W	4	30

$\chi^2 = 27.40$

$p < .001$

Σ of $\chi^2 = 118.9$, with 10 _df_, $p < .001$

Contingency tables for C-group subjects
The rows represent right ("R")and wrong ("W") choices for each subject. Columns represent right and wrong spontaneous paraphrases.

Table 11

S11

	R	W
R	101	0
W	3	1

S13

	R	W
R	106	5
W	5	1

S12

	R	W
R	87	9
W	0	2

S14

	R	W
R	65	1
W	4	4

S15

	R	W
R	105	0
W	2	1

Contingency Tables for A-Group Subjects

that subjects simply "know nothing" and are guessing, under this condition; quite the contrary, their errors (as we label them) seem to be systematic features of their organization of the task we set.

We now ask what happens when the subjects originally generated an error other than the modal error, and now have to choose between the correct paraphrase (which they did not generate) and an incorrect paraphrase which they also did not generate. Now the false alternative is someone else's error. Under these conditions, subjects are more likely to opt for the correct response, as Table 12 shows. For stimulus items on which a subject made a non-modal error in Experiment 1, the proportions of correct choices in Experiment 2 were generally over .5. The corresponding figures for stimulus items on which the subjects made a modal error in Experiment 1 were generally under .5. Every C-group subject shows a difference in the same direction.[1] Again, the A-group seems to show the same effect, but the numbers on which proportions for the A-group were calculated are ludicrously small. Thus the subject prefers his own response, right or wrong by our own criteria, and tends to our own definition of what is correct only when his own preferred response is not available to him.

---

[1] That this result is highly significant needs no statistical proof.

Table 12

% Correct Choice

Group	Subject	Where subject had generated the modal error	Where subject had generated a non-modal error
C	S1	35.4	48.7
C	S2	52.9	68.9
C	S3	49.2	65.0
C	S4	32.6	49.2
C	S5	36.0	64.7
C	S6	37.9	66.0
C	S7	16.1	52.0
C	S8	40.0	54.7
C	S9	50.0	56.7
C	S10	34.1	57.5
A	S11	0.0	100.0
A	S12	81.8	78.3
A	S13	83.3	90.0
A	S14	14.3	100.0
A	S15	0.0	50.0

3) Distribution of errors in the forced-choice situation: Because there were many new subjects in this condition, scores on the generating task were rescored for all subjects to see whether the results described for Experiment 1 were repeated with the new samples. Some refinement of the scoring technique was also attempted. The pattern of results was very much the same, so they will not be described here again. Appendix B explains the rescoring and gives the major results of this replication.

4. Experiment 3:

I mentioned briefly when introducing Experiment 2 (p.123) my conviction that the differences obtained in these experiments were not dialect-group distinctions. This conviction was based on three facts: (1) the secretaries have, in general, the same regional pronunciation (mid-atlantic urban) as the graduate students, by my observation; (2) to my knowledge, the distinctive stress features of compounds are very similar across the United States, perhaps excepting the Southern dialects; (3) the same syntactic structures were difficult for all populations tested, i.e. there was no interaction, in the spontaneous generation task, between group-membership and phrase-category. Nevertheless, in view of the nonoverlapping scores of A-group and C-group members, and the undeniable fact that there must exist secretaries who are more verbal, or more intelligent, than some graduate students, it seemed sensible to look for such secretaries and examine the structure of their responses. If we found secretaries who were as consistent as graduate students, but whose error-type preferences turned out to be similar to those of secretaries, we might conclude that there existed a dialect difference, at least in the mode of relating semigrammatical sentences to the underlying grammar.

At this writing, we have been able to find only two secretaries whose overall generation scores were at all close to those of the graduate students, although there was still no overlap: S16 had 55 errors, S17 had 47, while the maximum number of graduate-student errors was 34. Still, these two subjects err far less often than the thirty other secretaries we have so far tested. Table 13 shows the distribution of these subjects' errors by error-type; these distributions resemble A-group distributions (see Table 5) much more closely than they do C-group distributions.

In the forced-choice situation, these subjects again resemble the graduate students: S1 had 22 wrong choices; S2 had 19.

		Error Type:				
		Order	Stress	Chaos	Grammar	All
S16	No.	38	1	7	9	55
	%	69.1	1.9	12.7	16.3	
S17	No.	35	2	8	2	47
	%	74.6	4.2	17.0	4.2	

Table 13

Errors and distribution of error-types (in percentages) for two high-scoring secretaries.

Section IV: Discussion

There are three major findings of the experimental studies: (1) there are very large differences in the behavior of the three populations sampled in providing and recognizing paraphrases; (2) for all individuals, the same kinds of phrase-categories lead to difficulty in providing a paraphrase; and (3) the kinds of error 'preferred' differ for the various phrase-categories and for the three populations. In Section A of the discussion, I take up these topics in turn. Part 1 (Individual Differences) discusses the differences among people in their linguistic organization; part 2 (Grammaticalness and Paraphrasability) deals with the effect of syntactic structure in determing the nature of a paraphrase; part 3 (Semigrammaticalness and Error-Type) discusses modes of coping with linguistic material perceived as structurally deviant.

In Section B of the discussion, I take up the question of the relevance of this work for linguistic research. First the methodological implications of the results for linguists are discussed; second, the relevance of the work to the study of creativity and innovation in language use.

A. The Perception of Paraphrastic Relations

1. Individual Differences: Inescapably, these populations differ intheir responses to the experimental questions. Graduate students paraphrase compounds differently than High-School graduates.

a. But are there really individual differences? We have seen that the generative grammarians suppose that individuals are equally competent in linguistic matters. Since individual differences seem to exist, we must consider lines of argument by which these findings could be explained away, leaving the theory of equal competence intact. The first kind of argument would be that the experiment itself was at fault, i.e., that the apparent linguistic differences among the subjects are artifacts of the experimental situation. If the differences are facts, not artifacts, then the second position of the grammarians seems to be (cf. p. 43) that the experiment must therefore have tested performance, not competence. I consider both these matters here, not only because my results seem to be at variance with an influential theoretical position, but because the same questions ought to be answered about any experimental situation.

1) Fact or artifact? Let us ask once again: are the results an artifact of the experimental situation? Consider two possible artifacts that have some a priori plausibility:

(a) Perhaps the compounds were rather difficult to hear and to process, because of the absence of sentential context and determiners. This interpretation is not too likely. To begin with, it is virtually impossible that there were acoustic problems. Two judges reliably interpreted the stress-relations of the stimulus tapes before the experimentation began. Furthermore, there is clear evidence that the A-group subjects had no problems of this sort: they hear the stress and the separate words of the stimulus correctly in at least 35 of every 36 instances.[1] There is no reason to suppose that C-group subjects differ in their hearing from A-group subjects. I might add that, apart from errors of grammar, the paraphrases provided by the C-group do not seem to differ in content or ingenuity from those provided by the A-group: what differs is the match of phrase to paraphrase. There is then no reason to suppose that some part or attribute of the stimulus was not detected by the subjects, who then somehow had to fabricate that which they had missed. What differed from group to group was the decision of what to process or how to process it or the nature of the processing itself.

---

[1] Stress and chaos errors may result from mishearing of the stress, or they may involve an error in processing rather than encoding. However, all correct responses, and errors of other types imply (except for possible correct 'guesses' here and there) a correct hearing of the stress-relations. Thus if we consider all errors of stress and chaos to be encoding errors, we get the maximum number of mishearings of stress that could have occurred. On this basis, we conclude that A-group subjects mishear the stress no more than one time in thirty-six. Misreports of individual words occur in less than .001 of trials.

(b) Perhaps the subject did not understand the instructions. To be sure, the 'free' and 'forced-choice' situations produce essentially the same results, but after all, the instructions for both tasks were almost alike, so this point is not conclusive.

We can rule out this interpretation with fair assurance. The internal pattern of the results makes it certain that subjects understood well enough to behave systematically: (1) for all subjects, and for all stimuli, fewer than one response in thirty fails to elicit the syntactic format (a nominal or adjectival phrase) required for a correct response; (2) such format errors (errors of grammar) as did occur were almost totally limited to compounds with verb X-terms, for which there is often no conceivable grammatical interpretation; (3) for all subjects, the same stimulus types account for the same proportion of errors (e.g., there is no interaction between group-membership and phrase-category). All of these findings suggest that all subjects were responding to the same characteristics of the same stimuli.

There is a final variant of this hypothesis: perhaps the situation is so artificial that only the 'test-wise' individual can credit the kind of answer he is called upon to give. Perhaps it takes a sort of perverse sophistication to suppose that the right answer (to any question) could be 'one who blackens housebirds'. The results of the forced-choice study do much to remove this possibility, for here the subject has only a choice between eccentricities; yet he does not choose haphazardly between them.

It would appear, then, that the results reflect genuine group differences rather than artifacts of the experimental situation. In a way, this is not too surprising: individual differences in response to (apparently) linguistic tasks have been found before, at least where the tasks are not too simple (cf. Sleator and Maclay; Hill). An exception is the study by Livant (op. cit.) who found no variation at all among the three subjects he asked to paraphrase compound-nouns. Considering the size and reliability of our group effects, I can only conclude that Livant's three subjects were more linguistically sophisticated than the subjects in our A-group. If so, his study hardly constitutes a fair test of linguistic uniformity.

2) Performance or competence? These data make it impossible to deny that there are individual differences in the performance of linguistic tasks; Katz' assertion (cf. p.35) to the contrary is thus disposed of. But then there are few linguists who would not have granted

performance inequities from the beginning. To most of the linguists whose work is here relevant, the issue is not whether people differ in their performance, but whether such differences reflect any underlying differences in their linguistic competence.

It is obvious that the student of language initially confronts certain facts of performance; competence can only be inferred. On what basis can such an inference be drawn?

In order to make the distinction, the transformational linguist provides criteria of competence that are independent of such obvious facts about performance as the utterances a speaker happens to produce. Thus if a speaker can distinguish between a well-formed sentence and a deviant sentence, if he can recognize constructional homonymity, etc., we can infer a kind of organization not necessarily reflected in spontaneous speech.[1] Yet these procedures (so happily tailored to the special talents of professional linguists) cannot be the only ones allowed. Since at the heart of the transformational hypothesis is a description of the relations among sentences or sentence-types, it would seem that the ability to provide paraphrases (or at least to distinguish paraphrases from nonparaphrases) would be the most direct and compelling evidence for this theory. But if the theory of equal competence is correct, it should apply to secretaries no less than to graduate students. If, as Chomsky asserts, "each normal human has developed for himself a thorough competence in his native language", then in principle at least we should be able to extract this evidence from anybody. Necessarily I have asked subjects to 'perform' because I could not ask them to 'be competent.' Their paraphrastic performances varied widely. I now examine the Chomskian supposition that such enormous differences in performance are compatible with the hypothesis of equal-competence.

---

[1] I have previously noted that there have as yet been no general tests of competence -- for various populations -- using these procedures.

Performance deficits might reflect some kind of statistical noise in the system: random errors due to lapses of attention, lowered motivation, and the like; or they could be attributed to a failure to attend to the salient syntactic features of the task. I do not believe that these factors adequately account for the failure of some of our subjects to cope with the task presented to them.

There is no evidence from the results that C-group subjects were less reliable or systematic in their responses than A-group subjects. In the forced-choice situation, C-group subjects were very likely to choose a correct paraphrase if they had earlier generated one, and were very likely to choose an incorrect paraphrase if they had earlier generated it. Thus they do not make random errors: they choose systematically, but differently, from A-group subjects. Furthermore, sheer difference in the overall tendency to err will not account for the differences among groups. When we equate statistically for differing baselines for error, significant differences in the kind of error preferred by the various populations are not eliminated. It is thus difficult to argue that obtained differences between the groups represent merely different aptitudes in responding stably in test-situations.

It might be supposed that the C-group subject is more likely to be misled by semantic features of the situation that are essentially extraneous to the syntactic problem. The instructions for both experiments ask the subjects to disregard semantic anomalies ("many of these compounds may seem odd or silly to you..."), but this does not mean that all of the subjects were able to oblige us by disregarding them. One could suppose that, while the same underlying grammatical organization holds for all subjects, some of them are more prone to semantic seduction, more likely to desert the syntactic Super-Ego for the semantical Id. This supposition cannot be maintained in the face of the experimental findings.

In the first place, there is the evidence provided by the enormous effect of 'unity', the tendency to perceive a semantically plausible subunit even if the stress feature of the compound is thereby rendered syntactically deviant (e.g. *black*^{1}*house*^{3}*bird*^{2} is harder to paraphrase than *black*^{2}*house*^{1}*bird*^{3}). Here surely is an instance of semantic seduction. But it is no more apparent in C-group subjects than in A-group subjects, once the overall differences in error-rate are adjusted: the unity factor affects all groups equally (see Figure 3).

Another indication that C-group subjects were no more prone to semantic interference than A-group subjects comes from the results for the stimuli with 'mixed' lexical assignment. The notion 'a bird-house made of stone' is no more extravagant than the notion 'a stone used to make a bird-house'. Both phrases require a belief in the possibility of stone bird-houses. Yet given the stimulus *bird*^{1}*house*^{3}*stone*^{2}, only the C-group is inclined to respond 'a bird-house made of stone'. This tendency surely turns on the essential 'prepositionalness' of the word *stone* and not on the semantic character of the phrase.[1]

---

[1] I have earlier noted that *stone* is not even a 'good' adjective, i.e. we do not say 'a *very* stone house'. If it is a noun, then, it should be thought of as post-positional, and *stone*^{2}*bird*^{1}*house*^{3} should seem at least as difficult as *bird*^{1}*house*^{3}*stone*^{2}.

3) The limits of competence: Except for members of the A-group we have found little evidence of a general ability to recognize transformationally related phrases. Although, for each phrase, we provided in the forced-choice situation a legitimate grammatical paraphrase, C-group subjects did very poorly in choosing the appropriate paraphrase from a pile of two.

Then if competence is competence at anything, what is competence competence at? I have quoted various linguists and philosophers who, defining competence as the 'tacit knowledge' of the speaker, project into each normal adult what comes to an effective procedure for saying, classifying, and understanding an indefinitely extendable set of new sentences. In part, this claim is based on the selfevident fact that speakers can create 'new' sentences, physically different from any they have heard, by the use of certain iterative procedures, such as coordinating conjunction. An unbounded set of grammatical sentences is thereby brought into existence, but the ability to employ such devices does not imply the ability to understand complex embedded constructions of the language, nor the ability to classify sentences, nor the ability to recognize sentential relations in a principled way. These latter skills differ from person to person.

Whether the differences among people are so large that they interfere with reading The Daily News, we cannot know from what has been done here, and cannot really know at all until the role of grammatical organization in linguistic behavior is assessed. The linguistic entities I have studied are reasonably complicated, but three-word compounds are not computational monstrosities, and they are created and used all around us. Context and knowledge of the world no doubt resolve most of the confusions that inhere in these structures when they are encountered in everyday life. Nevertheless, the results of these experiments cast doubt on the ability of less-educated (or less linguistically able) individuals to deal with syntactic complexity when there is no contextual recourse.

In sum, there is little doubt that speakers differ in how well they can paraphrase difficult speech fragments. The effect cannot be dismissed as an artifact, nor can it be explained away as a 'mere' performance deficit. It is hard to maintain the position that competence is equal for all speakers unless one finds a more restricted and genuinely meaningful definition of that term.

2. Grammaticalness and Paraphrasability

We have by now ascertained that subjects and populations differ considerably in their ability to deal with the experimental stimuli. However, we also find effects that point to qualitative similarities in the behavior of all of the subjects.

a. Phrase-category: Phrase-category is a highly significant factor in determining response-type. For every stimulus item in the test, there is strong consensus about an appropriate response among A-group subjects, and this consensus persists in a rambling dwindling fashion down into the C-group, for all but the most difficult phrase-categories. In this kind of situation, with these instructions, compounds quite generally elicit as paraphrases relative clauses with the underlying structure of their 'transformational sources'; that is, right-hand structures in the formulas of Section III are related to their corresponding left-hand structures in a very special way. These rather uniform effects of phrase-category on the structure of responses must reflect something of what we mean when we say that all of our subjects belong to the same linguistic community.

Thus the significant effect of phrase-category is one more demonstration of the grammatical organization of linguistic behavior. I have mentioned various other behavioral correlates of syntactic structure: certain sequences are uniformly judged to be grammatical (cf. Chomsky, 1957; but also Sleator and Maclay) are easy to repeat (Johnson) and to recall and hear in noise (Marks and Miller; Miller and Isard). We can now add that the linguistic structure of a phrase has a powerful effect on the kind of paraphrase it will elicit without context, and often in spite of semantic or even syntactic deviance.

b. Phrase-category and the tendency to err: Some phrase-categories are significantly more difficult to paraphrase than others (see Table 4), i.e., they elicit fewer identical paraphrases, and indeed, elicit certain paraphrases that we reject. I have shown that the rank-order difficulty of the various phrase-categories is approximately the same for all populations, i.e. there is no interaction between group-membership and phrase-category measured against mean-error.

Discussed below are two possible kinds of explanation for the fact that certain phrase-categories are more difficult to paraphrase than others for all populations: (1) there may be differences in the grammatical status of the phrase-categories, and (2) there may be differences in the difficulty of the various types arising from the methods subjects use to encode linguistic material.

(1) Paraphrasability and the grammatical status of the compound: Table 7 shows that the lexical-class

of the test word has a highly significant effect on the tendency to err. Thus phrase-categories with noun test-words lead to fewer errors than phrase-categories with adjective and verb test-words. A possible explanation of this finding is that ungrammatical or semigrammatical sequences lead to much less uniformity in response-type: while the subject can deal with grammatical sequences by reference to a grammar, there is no such (or less of an) internalized 'semigrammar' that describes deviant sequences.

Referring to the grammatical sketch of Section III, we see that there are indeed many and confusing restrictions on the participation of verbs in compounds: verbs with certain kinds of complements are excluded; unaffixed forms of verbs are blocked in compounds by the existence of competing suffixed forms, etc. We know, in fact, that one of the verbs chosen (eat) results in uniformly unacceptable compounds.

However, if the grammatical status of the phrase-category is the factor determining its paraphrasability, we are left with some unanswered questions: The verb eat, which results in ungrammatical compounds, is associated with no more error than the verbs wash and kill, which almost uniformly form grammatical compounds. Further, compounds with post-nominal adjectives also lead to more error than compounds with noun test-words; yet all of these yield technically grammatical compounds.

It is clear that we cannot use paraphrasability as a criterion of grammaticalness, for that would involve us in a fruitless tautology. Criteria for grammaticalness have to do with systematic formal considerations. On the other hand, it is true that grammatical sentences (so defined by reference to precisely such formal criteria) very often are just those that are uniformly acceptable, paraphrasable, repeatable, recallable, etc. It would seem that such behavioral correlates of formal grammaticalness might give us some way of fixing a core of data for which the grammar is to account. However, when we adopt such an approach, we automatically build into grammatical description accounts of rather distant features of cognitive functioning (e.g. rules of grammar that were previously recursive are now to be limited by constraints on immediate memory). Furthermore, by accepting what is essentially Katz' position (cf. p. 47) we become incapable of describing aspects of linguistic behavior that are limited to certain individuals and to certain circumstances (we will discover as a trivial tautological consequence that people differ not at all in their perception of grammatical expressions in the language).

The failure to find that every compound we call grammatical is as easy to paraphrase as every other compound we call grammatical raises no obvious problems for grammatical description. But we still do not know why some compounds are easier to paraphrase than others. Consider, for example, the very large differences in difficulty between phrase categories with prenominal adjectives; e.g.:

black^{2}bird^{1}house^3
thin^{2}bird^{1}house^3 (C1)

and phrase-categories with postnominal adjectives, e.g.:

bird^{1}house^{3}black^2
bird^{1}house^{3}thin^2 (C2)

While C1 is easy to paraphrase, C2 is very hard. It scarcely needs showing that C1 represents a constructional type that occurs more frequently in everyday speech than C2, but why should that matter? If the subject 'knows' this fragment of syntax, and if the examples are not more unreasonably complex or nonsensical in the one case than in the other, why should subjects fail only with C2? Surely they have heard similar forms (midnight blue, peasoup green, ice cold, ruby red, sparking bright, dirt poor). Notice that even if in a grammar of English such transforms were 'generated later' or were 'further removed from the kernel' (and we have no reason to say they are), we would have no way of accounting for the fact that a person who 'knew the whole grammar' would have trouble with instances of such a constructional type if they did not strain his memory or his credulity.

Thus apparently we cannot turn to internal matters of grammatical complexity to decide why some phrase-categories lead to more errors than others. An obvious interpretation of the results, however, is that certain features of the grammar are known less well to speakers than are others. Postnominal use of adjectives is a case in point (phrase-categories A2-, A3+, and A3-) and compounding with verbs is another (phrase-categories V1, V2, and V3). Put another way, rules [iii] and [viii] of the grammar are very well assimilated by speakers, while rules [iv], [v] and [ix] are not.

If we accept this hypothesis, we must look at the results in absolute rather than relative terms, as we would for most questions of expertise: A-group subjects are a bit shaky on these frills of the grammar, but C-group subjects hardly seem to know them at all. Thus we can also explain the persistence of C-group subjects in repeating their errors systematically in the forced-choice situation, when the correct alternative is presented to them.

To account for a host of phrases constantly coming into the language in a regular and, to many, a comprehensible fashion, a grammar of English must describe postnominal adjectival phrases. By admitting this 'productive' process, we do not say exactly how productive it is, and it is not easy within grammatical description to make this proviso. If we say: "this transformation has strong cooccurrence restraints" we merely beg the question, for we haven't any notion how to describe those constraints. Yet it is intuitively clear that C1 and C2 phrases are examples of syntactic types of wholly different status within the language: C1 is almost totally productive, over the whole adjective-noun range, while C2 is obscurely limited to ill-understood adjective subcategories.

In general terms we can admit that there are "frills of grammar" of restricted status that must be accepted technically as legitimate -- because they are, to some extent, productive -- but which may be less well assimilated by speakers, perhaps just because of their highly restricted domain. However, we may find some more specific difficulties with the phrase-categories that give rise to nonuniformity in paraphrasing when we consider the possible encoding problems a speaker faces when he hears these stimuli.

(2) Paraphrasability and the encoding process: Another kind of explanation for the differential paraphrasability of various 'grammatical' compounds lies in a description of the process of encoding, which may mirror the grammar only remotely.

Chomsky (1961) has once suggested that a generative grammar "supplemented with a hierarchy of categories" may be the kind of device that best describes 'levels of grammaticalness' of utterances. The general schema Chomsky proposed is this:

> Suppose that we have a grammar that generates an infinite set of utterances with structural descriptions. Let us call the units in terms of which these utterances are represented by the neutral term formatives...[e.g., words, morphemes]...Suppose, in addition, that we have an *m*-level hierarchy of categories of formatives....On each level, the categories are exhaustive in the sense that each formative belongs to at least one....We might also require that each level be a refinement of the preceding one, i.e., a classification into subcategories of the preceding level.

Thus, for example, a sentence like Golf plays John has a grammatical representation at some level in the hierarchy of categories (NVN), but fails to be represented at another level when the noun-class is subcategorized into animate and inanimate, and the verb-class into verbs that can take inanimate subjects and verbs that cannot ($^*N_{in}V_{an}N_{an}$). The category-level at which the sentence 'fails' is the 'degree' to which it is ungrammatical. Thus, in the process of understanding a sentence, there may be a stage at which the utterance is projected as a sequence of various-level categories.

Suppose, as we listen to a phrase, we develop certain "expectancies" about the sequence of categories at some level that will occur within the phrase; violations of these expectancies may increase the probability of an error even though there are other situations in which we would be able to deal with such violations.[1] Thus at some level of representation, after N, or NN, A is unlikely. After N, or NN, unaffixed V is similarly unlikely. This does not mean that these sequences of categories are not grammatical, only unexpected.

This interpretation of the special difficulty of compounds with postnominal adjectives and verbs is not altogether at variance with syntactic description, for the assumption of syntactic organization of linguistic material does not preclude the existence of parallel heuristic procedures and computational devices in the encoder that are probabilistic and serial in nature. While left-to-right encoding of sentences fails to account for many linguistic phenomena (cf. Fodor and Bever), it is still quite likely that serial organization has its own relevance in the encoding process.[1] Should this be so, then violations of certain expectancies will interfere with comprehensibility.

---

[1] Left-to-right parsing devices employing phrase-structure grammars incorporate some such notions of expectancy. Yngve (81) relates difficulty for the hearer to the degree of embeddedness of an element, i.e. to the number of left-branches that connect it to higher-level constituents. There have been some experimental studies of this hypothesis (e.g. Martin and Roberts, 57).

This interpretation of the results helps to explain why compounds with postnominal adjectives and verbs, although grammatical, often fail to be paraphrased correctly: they are not economically encoded, and thus error arises.

The same hypothesis, on a lower level of the 'hierarchy of categories', explains the unity factor, the tendency to hear the two constant-nouns of the stimuli as though they were a unit. This is the most powerful determinant of error, aside from group-membership, in the test. We have shown that the prior familiarity of bird^{1}house^3 as a lexical item would not account for this factor, since the effect for the 'unknown' item house^{1}bird^3 was essentially the same (see p. 115). Yet both these sequences of words obey the cooccurrence constraints of the language, and both are semantically plausible. It may well be that once morphological as well as syntactic constraints are met in a contiguous pair of nouns within a nominal phrase, the hearer gets set to hear them as a unit. Violation of this expectancy seems to interfere systematically with comprehensibility: disunified sequences of this sort are reorganized and understood (or misunderstood) as unified sequences. Errors with these disunified sequences are twice as frequent as with unified stimuli.

---

[1] Johnson (op. cit.) observes that the first few words of sentences are repeated more often than the last few in an experimental situation designed to produce some errors in recall and repetition. This is a significant and stable effect. However, a larger effect, for the same sentences, is the tendency for all words of a major constituent to be repeated if one of them is repeated, i.e., the transitional error probabilities for words in the sentence are not independent. Thus while a phrase-structure description accounts for Johnson's results in large measure, an effect describable as serial organization in memory is also discernible.

The relation of habit or mental set to comprehensibility is noticable in the everyday use of language; I believe that a descriptive mechanism of this sort will be required to explain many facets of the relation between language-user and formal grammar. Not every pun is understood. We would not expect the speaker to realize that his intent was unclear every time he said 'Time flies', nor would we expect the listener to recognize it. What we might expect, at least in a large number of cases, is that both encoder and decoder would be able to recognize the ambiguity under certain circumstances; this much is a reasonable prediction from the theory of grammar. Thus we would expect the ambiguity to be recognized more often if we said: "Here is an ambiguous sentence: 'Time flies'". And we would expect it to be recognizable more often if, in addition, we paraphrased it: "I mean you should time flies, you should keep the time of flies."

In the spontaneous-generation task, we possibly have made it difficult for the subject to notice such problems with these constructions, given that their normal expectancy, or set, would interfere with their noticing them. Thus by the time the subject heard house - bird he might have formed a rather tight set (1) for house^1 bird^3, and (2) for a nominal construction, i.e., a following noun.

In the forced-choice task, we tried to break or weaken this set by presenting the subject with an alternative paraphrase. Should one expect this method to break the set? The answer depends on the kinds of assumptions that are made about the encoding process. If the subject hears the phrase, encodes it, and then simply looks for his chosen version on the response-cards, the situation is no different from the spontaneous-generation task. If, on the other hand, the choice of paraphrases is still open to the listener when he reads the response cards (and he is allowed to read them even before he hears the stimulus), the set ought to be weakened. The overall finding (see Table 10) is that subjects in the forced-choice situation tend strongly to do whatever they did in the spontaneous-generation task. Whether this is because their knowledge of the grammar disallows one of the alternatives, or whether it is because their encoding procedures preclude a change of mind, cannot be decided from these experiments.

3. Semigrammaticalness and error type

Let us ask, in terms of the expectancy hypothesis, what courses of action the subject may pursue when his set is violated. If, after a pair of constant nouns, the subject hears an adjective or a verb, rather than the expected noun, he has essentially the following choices:

(1) recast the word as the expected noun, e.g.:

house^{1}bird^{3}thin^{2} - a cracker shaped like a house-bird;
a diet food used by house-birds

(2) find a rule different from the one expected to be applicable, e.g.:

- as thin as a housebird

(3) a combination of (1) and (2):

- an annual festivity at which everybody tries to thin their house-bird

(4) recast the phrase as though the expectancies had been met, i.e., make an error, e.g.:

- a housebird who is thin

I adopt here the simplifying assumption that the subject will provide a correct and grammatically defensible paraphrase whenever he can. Given that a subject makes an error (chooses a grammatically indefensible paraphrase) I believe that he has perceived the stimulus as somehow other-than-grammatical and is attempting to cope what he conceives to be a 'semi-grammatical' sequence. I temporarily leave out of account the question of whether this perceptual organization arises during the encoding process, or thereafter.

As the scoring technique implies, we have noticed that subjects make different errors on different occasions. Sometimes they behave as though they had heard a stress pattern different from the one provided (E_s) and on other occasions they act as though they had heard the words in an order different from the actual one (E_o). If a subject 'commits' both these errors simultaneously, we say he has committed an error of chaos (E_{ch}). Further, there are other occasions on which the subject recasts the phrase (contrary to all sorts of grammatical cues, let alone the instructions for the task) as some other structure entirely, most notably an imperative (E_g). If the subject cannot see through to the implied nominal or adjectival structure of the compound as given, it follows that he must treat the stimulus as semigrammatical and make one of these adjustments.

To understand why subjects make different errors in response to different stimulus types and why different subjects make different errors, we must examine in some detail the question of how a person copes with perceived semigrammaticalness. This question has been the subject of much linguistic inquiry. By an ungrammatical sentence, I mean a sentence that a grammar does not describe. But it is obvious that the situation is not as black and white as that: very many sequences fail to be grammatical only by a little, and many such sequences are readily comprehensible. Chomsky, as discussed earlier, proposed a method of assigning 'degrees of grammaticalness' to such sequences by reference to a finite hierarchy of categories that are related to sentences at various levels of representation. Paul Ziff (82) has proposed that a speaker is equipped with a small set of procedural rules, in addition to the rules of grammar, that enable him to 'find his way back' from the partly grammatical (semigrammatical) sentence to a sentence described in the grammar. The length or tortuousness of the route back is Ziff's measure of the ungrammaticalness of the original sequence. Jerrold Katz (42) takes strong exception to Ziff's model, partly because Ziff's analysis leaves open the question of how the speaker recognizes the grammatical sentence to which the ungrammatical one was ultimately related --

his question is how the speaker knows what routes to choose. The answer, perhaps, is that Ziff's description takes up where Chomsky's leaves off. But Katz' major objections turn on the failure of Chomsky's proposal and Ziff's to partition the set of all nonsentences into two mutually exclusive and collectively exhaustive sets: the set of semisentences (which are partially structured and thus comprehensible) and the set of nonsentences (incomprehensible strings of words without recognizable underlying structure). Katz' own model involves an underground semisentence grammar, complete with its own semiphrase-structure, semiderivations, semitransformational rules, 'traffic rules' that limit the number and kind of semiderivations that can be undertaken in place of derivations and, further, rules that relate the resulting semistrings to the strings generated by the grammar.

The details of these models are beyond the scope of this discussion. However, our data provide some empirical evidence of the paths people in fact travel when they perceive a sequence as semigrammatical. It seems plausible that such evidence can provide a basis on which to narrow down the choice among theoretical models, if psychological validity is to be required of the model chosen.

To explain these subjects' activities with deviant sequences, we can equip them with the hierarchy of categories notion, and a set of simple heuristic devices such as Ziff proposes:

[sg i] Class extension: Word-class membership can be extended, or subcategorizations ignored.

[sg ii] Permutation: Sequences which are ungrammatical in the order presented can be reordered.

[sg iii] Change of contour: Sequences which are ungrammatical with the stress features presented can be repronounced.

[sg iv] Addition and deletion of constants.

Notice that the devices suggested here, and by Ziff, are just those operations (permutation, deletion, etc.) that appear in the structure of the grammar itself. We can closely predict a subject's paraphrases by means of these semigrammar 'rules' if we know (a) the sequence of categories that is being treated as semigrammatical, and (b) the population to which the subject belongs.

a. Phrase-category and error-type: The strength of the association between error-type and phrase-category (Figure 2) establishes the fact that subjects behave systematically, even when they perceive the stimulus as a grammatically deviant one. The choice of error, or 'route back', depends very heavily on the phrase-category of the stimulus: extension of class membership [sg i] occurs for all phrase-categories, but errors of order [sg ii] are associated with post-nominal adjectives, errors of stress [sg iii] with nouns, and errors of grammar [sg iv] with verbs, particularly in first-position. Thus the choice of response depends on how the phrase fails to be grammatical.

Generally speaking, subjects seem to isolate the anomaly, in these stimuli, by viewing the phrase as a sequence of elements in a phrase-structure representation, much as Chomsky suggests. Given that the sequence fails at this level, the subject makes selective changes, much in the way Ziff proposes.

For the kinds of stimuli we used, subjects[1] seem to prefer to find a higher-level of phrase structure at which the sequence has a grammatical representation, i.e. they ignore violated subcategorization rules [sg i]. We have scored such extensions as correct responses. Thus, overall, subjects 'correctly' paraphrase over half of the stimuli even though most stimuli violate cooccurrence restrictions or subcategorization rules.

---

[1] We have suggested earlier that [sg i] is almost totally limited to group-A subjects. When we pool the data for the various populations, [sg i] is a very frequent device. We consider the population restriction in the next section.

When subjects do not make use of this procedure either because they cannot, or because it does not work in the particular instance, they prefer to permute the sequence [sg ii] until it can be represented correctly at a high phrase-structure level. Thus errors of order are the most frequent for all populations.

When the test-word is an adjective, rearrangements of the elements, in this situation, will result [sg ii] in a sequence representable as a noun-phrase derivable with the 'well assimilated' rules [iii] and [viii]. When the test-word is a noun, rearrangement will help only if the resultant sequence now obeys cooccurrence restructions that were previously violated. The same result can often be achieved by a shift in stress [sg iii], and this route is very often selected when the sequence includes a noun test-word, i.e., disunified sequences are unified.

In a certain number of cases, subjects' responses can be characterized only by assuming that they applied both [sg ii] and [sg iii] to the same sequence. We have called such responses errors of chaos, since they seem to play so fast and loose with the stimuli as presented.

When the test-word is a verb, rearrangement of the elements or shift in stress, or both, will very often fail to affect the semigrammatical status of the sequence. Under this circumstance, some subjects seem willing to conclude that certain 'function words' and transformational constants may be missing from an imperative sequence. This is rather surprising, since the deletion of function words occurs only under very restricted and easily specifiable circumstances (in headlines; when Tonto speaks; when babies babble). Nevertheless, with other paths blocked, some subjects seemingly resort to an insertion of elements that make it possible to view verb-containing sequences as imperative sentences [sg iv].

Then adjective test-words are associated with [sg ii], nouns with [sg iii], and verbs with [sg iv]. Errors of chaos [sg ii + sg iii] are more uniformly distributed, although they are expectably more frequent for adjective test-words.[1] The use of [sg i] to extend word-class membership does not show up as an error in our scoring. The weight of evidence necessary in each instance to decide whether a subcategorization rule was violated (thus necessitating [sg i]) precluded reliable judgments by the scorers. We can only estimate the use of [sg i] by examining clear instances in which stimuli violated cooccurrence restrictions, and from our knowledge of the number of other kinds of error.

---

[1] The distribution of E_s, E_o, and E_{ch} is distorted slightly by the initial scoring technique, in which E_g [sg iv] superceded others, i.e. an error that involved E_g and E_o was scored as E_g.

b. Group-membership and the response to semigrammaticalness: Although error-type is related to phrase-category independently of group-membership, error-type also interacts with group-membership. That is, different populations prefer different kinds of error.

In the previous section, I noted that from these data and these scoring techniques one cannot decide at what point the subject decided that he was dealing with a semigrammatical sequence: we have had to score [sg i] corrections as correct responses. At any rate, inspection of the responses shows [sg i] to be largely restricted to A-group subjects, for clear instances. Inspection of the results for lexically ambiguous stimuli (Table 7) also shows a very large element of classificational rigidity for non-A-group subjects; this is another indication that C-group subjects are unlikely to use [sg i].

The massive relation between group-membership and the tendency to err (Figure 1) indicates that the point at which subjects decide to use the other [sg] rules differs significantly for different populations. The overall assumption is that the point at which subjects give up on the stimulus as a grammatical one and turn to [sg] rules differs significantly. Subjects also differ in the kind of error they then choose to make.

I have argued that errors of grammar [sg iv] do violence to the overt nominal structure of the sequence: they imply a set of missing determiners, affixes, etc., in a fairly arbitrary fashion. This being so, the use of [sg iv] - that is, an error of grammar - must be considered a serious error. One might also argue plausibly that the use of two [sg] rules is somehow worse and more serious than the use of one [sg] rule; thus errors of chaos can reasonably be considered 'serious' errors.

On these grounds we group together errors of stress and order as less serious errors, and errors of chaos and grammar as more serious errors. We find that, not only do A-group subjects err less often than C-group subjects, but they are significantly more prone to make less serious errors when they err at all. C-group subjects contribute almost all the errors of chaos and grammar. When the data are converted to relative numbers, thus correcting for the difference in the overall tendency to err, this difference between the groups continues to hold (Table 5). These differences are enhanced when a more refined scoring technique is adopted (see Appendix B).

The use of more serious errors by C-groups subjects for the same stimuli for which A-group subjects chose less serious errors implies either: (1) the C-group sujbect sees the stimulus as 'less grammatical' and so his 'route back' has to be more tortuous, or (2) the C-group subject is not economical in his use of [sg] rules, or fails to recognize that some [sg] routes are preferable to others.

We have already shown that, on an absolute basis, C-group subjects are significantly more prone to err than A-group subjects: they fail to perceive as many of the stimuli as grammatical, and they fail as well to make use of the acceptable [sg i] in case the stimulus is ungrammatical. We now see that they are also more prone to the serious [sg] alternatives than are the A-group members. From all these findings, we conclude that the paraphrases of the C-group are not merely different from those of the A-group, but qualitatively inferior.

I have not begun to suggest a formal description of the fact that people cope with semigrammaticalness systematically; I have here proposed one kind of technique for eliciting behavior that must constrain any such theory. The 'rules' or procedures postulated here are contextless, and perhaps specific to the particular set of stimuli I have chosen, though they are very similar to those proposed by Ziff on theoretical grounds. The results of the experiment seem to be a display of the kinds of systematic approach to semigrammatical material presupposed by both Ziff and Chomsky.

Katz might very rightly argue that the loose and vague collection of activities characterized by these [sg] rules is hardly the starting point for 'a theory of semsentences' that will explain what to do under what circumstances, which way to go first, how far to go, which sequences will be tractable and which not, why everybody understands the same semisentences in the same ways, why everybody distinguishes semisentences from nonsentences more or less identically.

However, there is really no obligation to meet the requirements of Katz' theory because it appears that people do different things under the same circumstances, go different ways first, go to different extents, find different sequences tractable, understand the same semisentences in different ways, and distinguish semisentences from nonsentences wholly differently.

The need for Katz' theory hangs on his belief that "Semisentences are comprehensible to each speaker according only to his linguistic abilities" (42) and on his already cited statement that "variation in performance with intelligence [in nonlinguistic tasks]...contrasts with the performance of speakers with respect to some purely linguistic skill, where no significant individual differences are found."

I believe on the basis of these experimental results that Katz' elaborate semigrammar may be too much apparatus with which to encumber the already overburdened normal speaker.

B. Some more general comments

1. Linguistic method

a. The Ideal Speaker-Hearer: Transformational linguists have ordinarily worked with sophisticated informants, choosing among them until they find one with 'a good ear'. To most linguists, this procedure needs little justification: the desirable subject is one who by some previous experience or innate talent comes equipped to answer questions of well-formedness reliably. On the other hand, this selection bias has sometimes been cause for dismay among critics (usually psychologists) more sensitive to matters of representative sampling who could not help asking what the population was to which the sample results were to apply.

A very similar issue has long been debated by psychologists concerned with problems in perception and psychophysics. In this case, the point is to relate certain psychological qualities (e.g., hue, form, perceived depth) to the physical attributes of the stimulus. The classical psychophysicists argued that normal individuals do not differ in the structure of their perceptual systems but do differ in the stability of their reports. Their solution was to train observers until their judgment became solid, and they assumed that the training did not 'change anything' about the character of these percepts themselves. The trained subject thus was supposed to be better able to get to the 'pure' percept, for his report would be uncontaminated by many of the random and systematic distortions his training had taught him to disregard. Whether this procedure is appropriate for psychologists is for them to decide. What is relevant here is the striking similarity between assumptions that are explicit in classical psychophysics and implicit in linguistics. Consider the assumptions necessarily made by a linguist who avoids subjects to whom the word 'grammatical' is obscure or provocative.

It is obvious that the method of asking only oneself (and perhaps for good measure one or two linguist friends) about linguistic questions is justifiable if and only if all normal persons are alike in their fundamental approach to grammar. However, we did find massive individual differences, both in the ability to paraphrase and in the way semigrammatical material is paraphrased. The notion that all normal individuals perform equivalently in linguistic tasks (which I take to mean, e.g., paraphrasing, classifying, punning, writing poetry) was not supported. Given these results, one may well suspect that grammars written by linguists may lack descriptive adequacy for normal populations: the linguists' informants are A-group members (that is, latent grammar-writers). The linguist has succeeded in setting up a mirror and has described himself.

If people differ, the failure to sample the linguistic community cannot but lead to a biased view of how people at large deal with language. The belief in equal competence is very possibly the upshot of assuming that linguists and philosophers provide a representative sample of the population: Speakers of English. Until we account for the fact that secretaries seem to be different in their linguistic competence from graduate students, we must reserve judgment on the hypothesis of equal competence -- as well as on all the nativist ramifications and neurophysiological suppositions that hang on it.

b. Data for a theory of language: There is no disagreement about the fact that linguistic inquiry begins with a list of obtained utterances. Hardly anyone seriously denies that this list must be accompanied by reports that at least give us the speakers' opinions about whether these utterances are reasonable representatives of the language under study, i.e., the utterances should not be those of an aphasic or a madman or a foreigner, nor utterances that were interrupted or poetic or humorously intended. Both the utterances and the introspective reports are facts about the performance of speakers, and from these we can begin to infer certain underlying abilities of speakers that may account for these performances in a systematic and illuminating way.

The utterances that people produce and their opinions about the legitimacy of these utterances already limit the kind of description we can adopt concerning linguistic organization. However, we can obviously choose more sensibly among models when we add to the stock of performances that we examine. Thus, taken together, the findings that people can create novel sentences, carry out paraphrasing, classifying and parsing tasks, and recognize constructional homonymity (even if they can do none of these things perfectly, or at all times) become in sum very powerful evidence for something like an internalized generative grammar. The 'scientific' assumption of certain empirical linguists that most of these kinds of data are irrelevant or unreal, that spontaneous utterances have some special status among performance types, hardly deserves serious consideration. I have discussed at great length, however, the problem of confusing these performances with the underlying abilities we set out to describe: failure to perform may not reflect the absence of an ability. This problem is not unique for questions of language. Psychologists have also pointed out (see in this connection Hochberg (34)) that the elicitation of introspective reports in such a way that the reliability of the interpretation may be evaluated requires varied and sophisticated techniques, convergent procedures, and internal validation. I am arguing in favor of such a pluralistic approach in linguistics.

The development of a systematic body of data by precise and sophisticated procedures seems to me to be a central step in the construction of an explanatory linguistic theory. The empiricist linguists never succeeded in learning how to collect usable data from human subjects. The generative grammarians, on the other hand, have simply not agreed that there is the need:

> The social and behavioral sciences provide ample evidence that objectivity can be pursued with little consequent gain in insight and understanding...At a given stage of investigation, one... must ask whether or to what extent a wider range and more exact description of phenomena is relevant to solving the problems that he faces. In linguistics, it seems to me that sharpening the data by more objective tests is a matter of small importance to the problems at hand....many questions that can realistically and significantly be formulated today do not demand evidence of a kind that is unavailable or unattainable without significant improvements in objectivity of experimental technique.
>
> (Chomsky, 1965)

Is there a current need for more objective data? I think there is, because I think that the generative grammarians have begun to draw conclusions from the data at hand, and the theory in hand, that could be justified only were some extensive further empirical confirmation provided. I believe that generative grammarians have been too quick to throw away disconfirmatory evidence simply by pointing to the performance-competence distinction. After all, failure to perform may sometimes be an indication of the lack of an ability, or a limitation on an ability. This is the other side of the coin.

It seems to me a strategic error to assume that what is in Chomsky must therefore be in his butcher; and that what is in the grammar must therefore be in the mind of the user. Perhaps grammarians leave too much out of account the fact that there are formal conditions on grammars such as internal consistency and simplicity that need not be requirements on the user. When individuals seem to differ, or fail, in displaying competence in apparently linguistic activities, we cannot categorically conclude that the test must therefore have been an irrelevant 'operational procedure', nor that the finding of difference itself is proof that whatever was tested lies outside the province of linguists (cf. Katz, p.47). Individual difference and limitations on capacity enter the province of linguistics at least at the point when linguists begin to argue from their formalism to the nature of the human mind. They enter well before that if questions of style, fluency, and articulateness in language use are linguistic questions.

c. The relevance to linguists of current empirical studies: Empirical predictions from the theory of grammar seem to be categorically confirmed when the grosser aspects of linguistic organization are studied. For example, Marks and Miller (56) and Miller and Isard (62) have demonstrated convincingly that speakers of English perceive some simple well-formed sentences wholly differently from simple non-well-formed sentences. Bever and Fodor (21) and Johnson (39) have shown that people perceive some high-level phrase-structural relations as well, and there do not seem to be any significant individual differences in responses to such tests. When the simple noniterative transformations are studied (e.g. Savin and Perchonock (70)), the result again seems to be in accordance with the predictions of linguistic theory, and free of individual difference.

However, when the response to more complex linguistic stimuli is examined (as here, and also cf. McNeill (55) and Blementhal (5)) or when further skills and intuitions related to grammatical organization are tested (as here, the ability to paraphrase, and also cf. Maclay and Sleator (52), the ability to classify sentences), individual differences in response-type begin to be found. And just as this more complex knowledge seems better developed in some individuals than in others, it seems to be better developed for certain fragments of the grammar than for others.

Thus while the psychological reality of phrase-structure grammars has been amply attested, the perception of transformational structure -- apart from the simple noniterative sentence-to-sentence transformations -- is not as easy to display.

Is this a technical accident? It is possible that the procedures involved in eliciting confirmatory evidence are here harder to develop. It is also possible that there is something real in the 'smear' or error-variance of these data. A forthcoming article by Martin Braine (8) describes work on language acquisition suggesting that phrase-structural organization of the language ontogenetically precedes transformational organization. Perhaps this reworking and resynthesis of the matter of language occurs in a different degree in different individuals. Perhaps, also, it fails to take place at all for certain kinds of constructions that we, nevertheless, have formal reasons for describing with transformational rules in a grammar.

I imagine most adults have had the experience of suddenly recognizing that some compound they had always considered a simple lexical item had a recognizable substructure never before noticed. Thereafter, is the decrypted item not restructured in perception? It seems plausible that the restructuring from immature phrase-structure representations of linguistic material (in which, presumably, there are many kinds of 'simple' sentences, i.e., there are many primitive symbols 'S') into consolidating transformational representations takes place only for some of the material, and for some of the people. The hundreds of attested instances of certain postnominal uses of adjectives (e.g., onyx black) may be insufficient to achieve rule [ix] in the mind of the speaker.

Empirical work of the sort cited here at least gives direction to such inquiry, and it reveals details of phenomena that are not easy to come by intuitively or logically. These studies are interesting not because they test or affirm or make more objective generative grammars; they are interesting because they help us assess more realistically what generalizations about linguistic processes and the nature of the mind can be drawn from formal grammars.

2. Innovation and creativity in language use: Generative grammar enables us to describe novelty or creativity in the use of language through the speaker's ability to make use of recursive procedures that are inherent in language structure. Only when we understand these laws or rules of language can we explain the everyday existence of forms like Volume feeding management success formula award, and their apparent comprehensibility. To attempt to explain

the behavior these entities imply without reference to a systematic grammar of the language is a self-defeating enterprise, although it has been tried often enough within linguistics and within psychology. As Postal (66) has put it:

> As if one were to suggest studying the behavior of a computer without knowing anything about its program, or the activity of chess players without having studied the rules of chess, or the performance of a symphony without considering the score.

At the same time, if we simply equate the speaker's behavior, or even his underlying competence or knowledge, with the rules of language themselves, we erase the distinction between speaker and system, between language and speech, as surely as do the linguists and psychologists who denied the distinction in the first place. Language is no more a projection of possible perceptual judgments than it is a projection of possibly obtainable utterances. It is fairly obvious that few, or no, individuals could on a single occasion paraphrase:

> the award given for succeeding in finding a formula for managing the feeding of people in large volumes

by its compound-nominal equivalent. The compound itself is the momument, no doubt, of a long entrepreneurial history, in which the establishment of the journal Volume Feeding Management was an epochal event. Since the Language, in this sense, is bigger than any of us, and since none of us, in this sense, is wholly competent to use it, it does little enough harm to recognize that among us there will be differences in the extent of our approach to the abstraction.

We have seen then that the abstract property of recursiveness in language structure explains certain creative activities of the speaker, although the Language itself is not limited by the limitations, whatever they are, on the speaker's creativity at any one moment. Yet there is a further distinction between the 'rules' of an abstraction like Lanugage and the rules of an abstraction like Chess:[1] should my daughter steal my queen or move her king two squares (these are by no means merely hypothetical events), the rules of chess are not thereby affected. Similarly, if the whole second-grade fails an arithmetic test, the structure of arithmetic is left unaltered.

This is not the case for language. If the second-grade persists in saying chess-wise, where we have always said in a chess-like fashion, that decision will eventually rebound upon the language itself.

---

[1] I leave aside here the question of 'conscious' vs. 'tacit' rule-following behavior.

The problem of the ways we learn and use what we know about language thus become an interesting linguistic question, for it is to a certain extent the question of the mutability of language structure. If there are aspects of English syntax which, by virtue of their own internal logic, or by virtue of their relations to other aspects of the language, or by virtue of the demands they make upon the mind of the user, are imperfectly learned, or learned differently by those who are less capable than others, this imperfect knowledge may give rise to a systematic set of utterances ("errors", to us) that are synthesized into new ("correct") rules by the succeeding generation.

Thus, I think, when we turn to a consideration of how people know what they know about linguistic structure, and how they cope with what they do not know (how they deal with perceived semigrammaticalness), we begin to approach another source of creativity and innovation, beyond recursivity, in language use. I believe this source of creativity is close to what is usually called analogic change. Poets make analogic extensions when it suits their purposes; so, in quite another fashion, do those whose lack of capability may force them to deal with nondeviant sequences as though they required such extension.

For the case of compound-nouns, we are fortunate in being able to chronicle the life-cycles of productive processes that have at least temporarily ceased to operate in English. For the disused compounding processes leave morphological fossils in the lexical stock of the language: we see cut-throat and pick-pocket where we would now 'construct' throat-cutter and pocket-picker. Indeed, what is productive in compounding seems to change in English at a relatively quick rate, if we compare these changes to innovations, let us say, in processes for sentence-negation or modes of affixing clauses to nouns. Could this be because of the inherent difficulties in processing or understanding compounds? When we consider how our subjects deal with difficult compounds, we seem to get a hint about their preferences in analogic extension, whereever they are forced to make them: thus the sequential pattern of the English compound is seemingly more often ignored, or ignorable, for the speaker, than its internal stress relations. This last, of course, is a hunch based on a small sample of very artificial data, and nothing really can be made of it: I offer it only as a suggestion for the line that research into such questions might profitably follow.

In sum, the Language seems to have an existence of its own, independent of the abilities and disabilities of those who bend it to their uses. Its creative quality derives in part from its recursive properties, but these very properties tax qualities of the speaker's mind and intention.

It is at this point that new forms are analogically created, again in a systematic way. Thus perhaps language may develop and change through constant and strenuous interaction between its own genius and the speaker's attempt to cope with it within the structure of his limited perceptions.

Appendix A

List of Stimuli

1. foot^1 bird^3 house^2
2. house^2 shut^1 bird^3
3. black^1 house^3 bird^2
4. bird^1 stone^3 house^2
5. house^1 wash^3 bird^2
6. bird^1 house^3 boot^2
7. glass^2 house^1 bird^3
8. house^2 bird^1 eat^3
9. house^1 bird^3 kill^2
10. bird^2 house^1 bright^3
11. thin^2 bird^1 house^3
12. bird^2 dry^1 house^3
13. kill^1 house^3 bird^2
14. bird^2 boot^1 house^3
15. house^2 glass^1 bird^3
16. house^2 bird^1 dry^3
17. bird^1 house^3 foot^2
18. house^1 bird^3 stone^2
19. wash^1 bird^3 house^2
20. shut^2 bird^1 house^3
21. eat^2 house^1 bird^3
22. bird^1 bright^3 house^2
23. house^1 black^3 bird^2
24. bird^2 house^1 thin^3
25. house^2 thin^1 bird^3
26. bird^2 eat^1 house^3
27. wash^1 house^3 bird^2
28. shut^2 house^1 bird^3
29. bird^1 house^3 dry^2
30. house^1 bird^3 black^2
31. house^1 foot^3 bird^2
32. bird^2 house^1 stone^3
33. bird^1 kill^3 house^2
34. bright^2 bird^1 house^3
35. house^2 bird^1 glass^3
36. boot^1 bird^3 house^2
37. stone^2 bird^1 house^3
38. dry^1 bird^3 house^2
39. house^2 bird^1 shut^3
40. house^1 bird^3 wash^2
41. foot^1 house^3 bird^2
42. house^1 boot^3 bird^2
43. house^2 bright^1 bird^3
44. bird^1 black^3 house^2
45. bird^2 glass^1 house^3
46. bird^2 house^1 kill^3
47. thin^2 house^1 bird^3
48. bird^1 house^3 eat^2

49. eat^1 bird^3 house^2
50. house^2 stone^1 bird^3
51. boot^1 house^3 bird^2
52. bright^2 house^1 bird^3
53. kill^2 bird^1 house^3
54. bird^1 wash^3 house^2
55. bird^2 shut^1 house^3
56. house^1 bird^3 foot^2
57. house^1 dry^3 bird^2
58. bird^2 house^1 black^3
59. house^2 bird^1 thin^3
60. bird^1 house^3 glass^2
61. bird^2 thin^1 house^3
62. bird^2 house^1 wash^3
63. glass^1 bird ^{3} house^2
64. stone^2 house^1 bird^3
65. bird^1 foot^3 house^2
66. bird^1 house^3 shut^2
67. house^1 eat^3 bird^2
68. black^2 bird^1 house^3
69. house^1 bird ^{3} boot^2
70. dry^1 house^3 bird^2
71. house^2 kill^1 bird^3
72. house^2 bird ^{1} bright^3
73. house^1 bird^3 dry^2
74. shut^1 bird^3 house^2
75. bird^2 bright^1 house^3
76. bird^2 house^1 foot^3
77. bird^1 boot^3 house^2
78. wash^2 bird^1 house^3
79. house^2 black^1 bird^3
80. kill^2 house^1 bird^3
81. bird^1 house^3 thin^2
82. house^2 bird^1 stone^3
83. house^1 glass^3 bird^2
84. eat^1 house^3 bird^2
85. house^1 shut^3 bird^2
86. black^2 house^1 bird^2
87. bird^2 house^1 boot^3
88. glass^1 house^3 bird^2
89. house^2 wash^1 bird^3
90. bird^2 stone^1 house^3
91. house^2 bird^1 kill^3
92. bird^1 house^3 bright^2
93. thin^1 bird^3 house^2
94. house^1 bird^3 eat^2
95. bird^1 dry^3 house^2
96. foot^2 bird^1 house^3
97. house^2 foot^1 bird^3
98. bird ^{1} eat^3 house^2
99. house^1 bird^3 glass^2
100. shut^1 house^3 bird^2
101. house^1 thin^3 bird^2
102. boot^2 bird^1 house^3

103. wash^2 house^1 bird^3
104. bird^2 house^1 dry^3
105. bird^1 house^3 stone^2
106. bird^2 kill^1 house^3
107. bright^1 bird^3 house^2
108. house^2 bird^1 black^3
109. foot^2 house^1 bird^3
110. house^2 boot^1 bird^3
111. house^1 bird^3 shut^2
112. bird^1 house^3 kill^2
113. thin^1 house^3 bird^2
114. stone^1 bird^3 house^2
115. bird^2 black^1 house^3
116. bird^2 house^1 eat^3
117. dry^2 bird^1 house^3
118. bird^1 glass^3 house^2
119. house^2 bird^1 wash^3
120. house^1 bright^3 bird^2
121. house^1 bird^3 thin^2
122. bird^2 house^1 glass^3
123. house^1 stone^3 bird^2
124. eat^2 bird^1 house^3
125. bird^1 shut^3 house^2
126. boot^2 house^1 bird^3
127. bird^1 house^3 black^2
128. bird^2 wash^1 house^3
129. house^2 bird^1 foot^3
130. kill^1 bird^3 house^2
131. house^2 dry^1 bird^3
132. bright^1 house^3 bird^2
133. bird^1 house^3 wash^2
134. glass^2 bird^1 house^3
135. house^1 kill^3 bird^2
136. house^2 bird^1 boot^3
137. bird^1 thin^3 house^2
138. dry^2 house^1 bird^3
139. house^1 bird^3 bright^2
140. bird^2 house^1 shut^3
141. black^1 bird^3 house^2
142. bird^2 foot^1 house^3
143. house^2 eat^1 bird^3
144. stone^1 house^3 bird^2

Appendix B

Distribution of errors in the generating task, and a new scoring technique.

Distribution of errors:

1. Rescoring the results: When describing the original scoring technique (P. 90), we mentioned the possibility of distinguishing errors that invert the elements of the 13 ("primary") compound (substituting bird-house for house-bird) and errors that invert the elements of the "secondary" compound (substituting black bird-house for bird-house black). This latter error seems intuitively less serious than the former. We now distinguish between a primary order error (O_p) and a secondary order error (O_s).

There was one further refinement we now tried to make in the scoring technique. The subjects sometimes gave sentential responses where relative clauses were called for; in all these cases, the response was scored as E_g, a grammatical error. But sometimes this error seemed to be trivial -- in cases where it was otherwise a sensible paraphrase, e.g. a subject might say "The bird-house was black" instead of "the bird-house that was black". In other cases, the error seemed more serious because it included, as well, a stress, order, or chaos error. In this version, we double scored sentential responses, so that we could separate those that were otherwise correct from those that included a further error.

Table I shows the distribution of error scores in generation using the old and the new techniques. It will be recalled that the population used in Experiment 1 (where we employ the original scoring technique) differs in part from that used in Experiment 2 (where we employ the refined scoring technique). It can be seen immediately that the results are very similar despite changes in the population and the scoring. The difference in distribution of error-types among the groups is slightly enhanced by the new scoring technique, for O_p distributes itself fairly evenly among the three population-groups, thus increasing the differences in the proportion of O_s errors among groups. Since grammar errors (E_g) very generally incorporated some further internal error, (generally an *order* or *chaos* error), the double scoring of results does not significantly decrease the absolute number of errors.[1]

Table II shows the distribution of error-type by lexical class for the populations of Experiment 2, using the new scoring. A comparison of this table with the bargraph of Figure 2 shows that these relations remain very much the same.

---

[1] Inspection of Table I does show a significant drop in the proportion of grammar errors for the A-group. The two A-group subjects who contributed these errors were not available for the second experiment, and no new A-group members made more than one or two such errors.

Table I

	Group					
	A		B		C	
	Scoring		Scoring		Scoring	
	old	new	old	new	old	new
Secondary Order		52.94		31.83		23.95
Primary Order		12.74		13.90		15.35
Order (combined):	56.6	65.68	38.1	45.73	26.6	39.30
Stress	24.8	23.52	31.0	31.39	34.0	30.07
Chaos	6.2	10.78	23.8	22.86	23.1	30.61
Grammar	12.4	3.92	7.1	4.03	16.3	17.56

Effects on error-type distribution of rescoring and some change in population (percentages sum to more than 100 where E_g is double-scored).

Table II

Lexical Class	O_s	O_p	S	Chaos	Grammar
N	24.88	11.96	38.75	25.40	6.2
A	31.26	17.63	28.25	22.84	4.0
V	32.37	10.79	18.28	38.54	28.85
AV	30.44	13.68	28.49	27.37	14.80
NA	21.11	18.51	41.48	18.88	2.22

Distribution of error-type for each lexical class using the new scoring and new population (percentages sum to more than 100 where E_g is double-scored).

Bibliography

1. Berko, J., "The child's learning of English morphology," in Saporta, S., ed., Psycholinguistics, Holt, Rinehart and Winston, N. Y., 1961.
2. Blackwell, H. R., and Schlosberg, H., "Octave generalization, pitch discrimination and loudness thresholds in the white rat," J. exp. Psychol., 33, 1943, 407-419.
3. Bloch, B., editor's note to Postal (66).
4. Bloomfield, L., Language, Henry Holt and Co., N. Y., 1933.
5. Blumenthal, Arthur L., "Observations with self-embedded sentences," Psychonom. Sci., 6.10, 1966, 453-454.
6. Bolinger, Dwight, "The atomization of meaning," Lg., 41.4, 1965, 555-573.
7. Braine, M. D. S., "The ontogeny of English phrase-structure: The first phase," Lg., 39, 1963, 1-13.
8. Braine, M. D. S., "The acquisition of language in infant and child," to appear in Reed, L., ed., The Learning of Language, National Council of Teachers of English.
9. Brown, R., "Linguistic determinism and the part of speech," J. Abnorm. soc. Psychol., 55, 1957, 1-13.
10. Brown, R., Words and Things, The Free Press, Glencoe, Ill., 1958.
11. Brown, R., and Bellugi, U., "Three processes in the child's acquisition of syntax," in Lenneberg, E. H., ed., New Directions in the Study of Language, The MIT Press, Cambridge, 1964.
12. Brown, R., and Berko, J., "Word association and the acquisition of grammar," Child Develpm., 31, 1960, 1-14.

13. Carroll, J. B., *Language and Thought*, Prentice-Hall, Englewood Cliffs, N. J., 1964.

14. Chomsky, Noam, *Syntactic Structures*, Mouton and Co., The Hague, 1957.

15. Chomsky, Noam, *Current Issues in Linguistic Theory*, Mouton and Co., The Hague, 1964a.

16. Chomsky, N., "Degrees of grammaticalness," in Fodor, J. A., and Katz, J. J., *eds*., *The Structure of Language*, Prentice-Hall, Englewood Cliffs, N. J., 1964b.

17. Chomsky, Noam, *Aspects of the Theory of Syntax*, The MIT Press, Cambridge, 1965.

18. Davis, K., "Final note on a case of extreme social isolation," *Amer. Jl. Sociol*., ***52***, 1947, 432-437.

19. Ervin, Susan M., "Imitation and structural change in children's language," in Lenneberg, E. H., *ed*., *New Directions in the Study of Language*, The MIT Press, Cambridge, 1964.

20. Fisher, Sir Ronald A., *Statistical Methods for Research Workers*, Hafner Publishing Co., N. Y., 1958.

21. Fodor, J. A., and Bever, T. G., "The psychological reality of linguistic segments," *J. verb. Learn. verb. Behav*., ***4***.5, 1965.

22. Fries, Charles C., *American English Grammar*, Appleton-Century Co., N. Y., 1940.

23. Gilbert, M., *Death Has Deep Roots*, Lancer Books, N. Y., 1964.

24. Gleitman, L. R., "A study of the acquisition of syntactic structure." Paper read at Linguistic Circle of New York, March, 1965a.

25. Gleitman, L. R., "Coordinating conjunctions in English," *Lg*., ***41***.2, 1965b, 260-293.

26. Goldstein, K., *Language and Language Disturbances*, Grune and Stratton, N. Y., 1948.

27. Greenough, J. B., and Kittredge, G. L., *Words and their Ways in English Speech*, Beacon Press, Boston, 1962.

28. Harris, Z. S., "Discourse Analysis," *Lg*., **28**, 1952, 18-23

29. Harris, Z. S., "Co-occurrence and transformation in linguistic structure," *Lg*., **33**, 1957, 293-340.

30. Harris, Z. S., Lectures on transformational linguistics, Univ. of Pa., Phila. Pa., 1963.

31. Hayes, C., *The Ape in our House*, Harper, N. Y., 1951.

32. Hill, A. A., "Grammaticality," *Word*, **17**, 1961.

33. Hiz, Henry, Personal communication. 1967.

34. Hochberg, J. E., "Perception: Toward the recovery of a definition," *Psychol. Rev.*, **63**, 1956, 400-405.

35. Hockett, C. F., *A Course in Modern Linguistics*, Macmillan, N. Y., 1958..

36. Hovland, C. I., and Weiss, W., "Transmission of information concerning concepts through positive and negative instances," *J. exp. Psychol*., **45**, 1953, 175-182.

37. Irwin, O. C., "Infant speech: effect of systematic reading of stories," *Jl. of Speech and Hearing Res*., **3**, 1960, 187-190.

38. Jakobson, R., and Halle, M., *Fundamentals of Language*, Mouton and Co., The Hague, 1956.

39. Johnson, Neil F., "The psychological reality of phrase-structure rules," *J. verb. Learn. verb. Behav*., **4**.6, 1965, 469-475.

40. Joos, Martin, *ed*., *Readings in Linguistics*, Washington, 1957.

41. Kahane, H., Kahane, R., and Saporta, S., "Development of verbal categories in child language," *IJAL*, **24**, 1958.

42. Katz, J. J., "Semi-sentences," in Fodor, J. A., and Datz, J. J., eds., The Structure of Language, Prentice-Hall, Englewood Cliffs, N. J., 1964.

43. Katz, J. J., and Fodor, J. A., "The structure of a semantic theory," Lg., 39, 1963, 170-210.

44. Kellogg, W. N., and Kellogg, L. A., The Ape and the Child, McGraw-Hill, N. Y., 1933.

45. Lees, R. B., The Grammar of English Nominalizations, Mouton and Co., The Hague, 1960.

46. Lenneberg, E. H., "A biological perspective of language," in Lenneberg, E. H., ed., New Directions in the Study of Language, The MIT Press, Cambridge, 1964a.

47. Lenneberg, E. H., "Speech as a motor skill with special reference to non-aphasic disorders," in Bellugi, U., and Brown, R., eds., The acquisition of language, Child Develpm. Monogr., 29, 1964b, 115-126.

48. Lieberman, P., "On the acoustic basis of the perception of intonation by linguists," Word, 21.1, 1965, 40-54.

49. Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, Houghton Mifflin Co., Riverside Press, Cambridge, Mass., 1963.

50. Livant, W. H., "Productive grammatical operations I - the noun-compounding of 5-year olds," Lg. Lrning., 1961.

51. Lorenz, K., King Solomon's Ring, Crowell; N. Y., 1952.

52. Maclay, H., and Sleator, M., "Responses to language: judgments of grammaticalness," IJAL, 26, 1960, 275-282.

53. McCarthy, D., "Language development in children," in Carmichael, L., ed., Manual of Child Psychology, 2nd ed., John Wiley and Sons, N. Y., 1954.

54. McMahon, L. E., "Grammatical analysis as a part of understanding a sentence," Ph.D. thesis, Harvard University, 1963.

55. McNeill, D., "Speaking of space," *Sci.*, **152**, 1966, 875-880.

56. Marks, L., and Miller, G. A., "The role of semantic and syntactic constraints in the memorization of English sentences," *J. verb. Learn. verb. Behav.*, **3**, 1964, 1-5.

57. Martin, E., and Roberts, K. H., "Grammatical factors in sentence retention," *J. verb. Learn. verb. Behav.*, **5**.3, 1966, 211-218.

58. Menyuk, P., "Alternation rules of children's grammar," *J. verb. Learn. verb. Behav.*, **2**, 1963, 346-351.

59. Miller, G. A., "Some psychological studies of grammar," *Amer. Psychol.*, **17**, 1962, 748-762.

60. Miller, G. A., "Language and psychology," in Lenneberg, E. H., ed., *New Directions in the Study of Language*, The MIT Press, Cambridge, 1964.

61. Miller, G. A., Galanter, E., Pribram, K. H., *Plans and the Structure of Behavior*, Holt, Rinehart and Winston, N. Y., 1960.

62. Miller, G. A., and Isard, S., "Some perceptual consequences of linguistic rules," *J. verb. Learn. verb. Behav.*, **2**.3, 1963, 217-228.

63. Moore, J. K., "Speech content of selected groups of orphanage and non-orphanage pre-school children," *J. Exp. Ed.*, **16**, 1953, 161-167.

64. Neisser, U., *Cognitive Psychology*, Appleton-Croft, in press.

65. Nida, E. A., *Morphology: The Descriptive Analysis of Words*, The University of Michigan Press, Ann Arbor, 1949.

66. Postal, P., Review of Dixon, R. M. W., *Linguistic Science and Logic*, *Lg.*, **42**.1, 1966, 84-93.

67. Putnam, H., "Some issues in the theory of grammar," in Jakobson, R., ed., Structure of Language and its Mathematical Aspects, Proceedings of the 12th Symposium in Applied Mathematics, Amer. Mathematical Soc., Providence, 1961.

68. Sapir, E., Language, Harcourt, Brace and Co., N. Y., 1921.

69. Sapir, E., Culture, Language and Personality, Mandelbaum, D. G., ed., Univ. of Cal. Press, Berkeley, 1957.

70. Savin, H. B., and Perchonock, E., "Grammatical structure and the immediate recall of English sentences," J. verb. Learn. verb. Behav., 4.5, 1965.

71. Schuell, H., and Jenkins, J. J., "The nature of language deficit in aphasia," in Saporta, S., ed., Psycholinguistics, Holt, Rinehart and Winston, N. Y., 1961.

72. Seuss, Dr., One Fish Two Fish, Red Fish Blue Fish, Beginner Books, N. Y., 1960.

73. Shipley, E. F., Smith, C. S., and Gleitman, L. R., "Free responses to verbal stimuli in one and two-year olds," Technical Report, 2, MH 07990, Eastern Pa. Psychiatric Institute, Phila. Pa., 1965, (mimeo).

74. Skinner, B. F., Verbal Behavior, Appleton-Century Crofts, N. Y., 1957.

75. Smith, C. S., "Two studies of the syntactic knowledge of young children." Paper read at Linguistics Colloquim, MIT, November, 1966.

76. Thorndike, E. L., and Lorge, I., The Teacher's Word Book of 30,000 Words, Teachers College, Columbia University, N. Y., 1963.

77. Wason, P. C., "The processing of positive and negative information," Quart. J. exp. Psychol, 11, 1959, 92-107.

78. Wason, P. C., "Response to affirmative and negative binary statements," Brit. J. Psychol., 52, 1961, 133-142.

79. Wason, P. C., "The contexts of plausible denial," *J. verb. Learn. verb. Behav.*, 4.1, 1965, 7-11.

80. Weir, R. H., *Language in the Crib*, Mouton and Co., The Hague, 1962.

81. Yngve, V. H., "A model and an hypothesis for language structure", *Proc. Amer. Phil. Soc.*, 104, 1960, 444-466.

82. Ziff, Paul, "On understanding 'understanding utterances'," in Fodor, J. A., and Katz, J. J., eds., *The Structure of Language*, Prentice-Hall, Englewood Cliffs, N. J., 1964.